B & Z LIBRARY

OHIO STATE UNIVERSITY
APR 24 1961
LIBRARY

Mathematical Models for use in Insect Pest Control

by

KENNETH E. F. WATT
Statistical Research and Services
Research Branch, Ottawa, Canada

THE CANADIAN ENTOMOLOGIST
Supplement 19
Accompanies Volume XCIII, 1961

Mathematical Models for use in Insect Pest Control

by

KENNETH E. F. WATT

Statistical Research and Services
Research Branch, Ottawa, Canada

THE CANADIAN ENTOMOLOGIST

Supplement 19

Accompanies Volume XCIII, 1961

CONTENTS

PAGE

1. Introduction 5

2. The Need for Mathematical Models of Pest Population Dynamics 7

3. An Approach to the Problem of Developing a Model 9

4. Development of an Inductive Model 16

5. Development of Mixed Deductive-Inductive Models 27

6. Methods of Using Models 45

7. The Field Program Dictated by the Model 51

8. Possible Criticisms of the Model Approach 55

References 57

Mathematical Models for use in Insect Pest Control

By KENNETH E. F. WATT
Statistical Research and Services
Research Branch, Ottawa, Canada

1. INTRODUCTION

No satisfactory comprehensive mathematical model has ever been developed to account for changes in the abundance of a specific natural population of insects or other pests. It is interesting for two reasons that such a model has never been developed. First, it would be very useful, as we shall show in the next section. Second, methods of constructing mathematical population models have been clear to some workers for over three decades (e.g. Thompson, 1924; Lotka, 1925; Volterra, 1926; Bailey, 1931; Kostitzin, 1939). Many other scientists, aware of models being developed by theoreticians, have repeatedly urged that data be collected and analyzed with a view to verifying the assumptions on which deductively-derived equations were based.

For example, in the foreword to Gause's (1934) book, Raymond Pearl stressed that the struggle for existence described by Darwin was a matter of population dynamics, then wrote as follows. "If ever an idea cried and begged for experimental testing and development, surely it was this one. Yet the whole array of experimental and statistical attempts in all these years to produce some significant new evidence about the nature and consequences of the struggle for existence is pitifully meager". It is noteworthy that Pearl didn't mention the most difficult problem of all, that of analyzing natural populations. The progress of work in population dynamics was correctly assessed by Leslie (1946) in a review of d'Ancona's (1942) book surveying population theory. Leslie noted that field studies verifying the mathematical reasoning occupied only 14 pages of d'Ancona's 341 page book, and that there was an essentially unbridged gap between the mathematician and the experimentalist. He concluded that the gap would remain until biological work in population dynamics was developed much further, and noted, "It is an amusing, and a far from unprofitable speculation to consider the way in which theoretical physics might have developed if only an occasional experiment had been done".

As of the present, a great deal more has been accomplished on the mathematical and experimental approaches to population dynamics, but work on populations in their natural surroundings is still relatively unevolved, with certain exceptions. Fish population dynamics is becoming a highly sophisticated science (e.g. Beverton and Holt, 1957; Bishop, 1959; Paloheimo, 1958; Ricker, 1958; Schaefer, 1957). Sampling, and matters relating to life tables are being carefully analyzed by some entomologists (e.g. LeRoux and Reimer, 1959; Morris, 1955, 1957, 1960; Morris and Miller, 1954).

However, there is still no model anywhere in the literature of population dynamics that comes remotely close to simulating the following features of an insect population's environment.

1. There exist a great array of factors relevant to population dynamics of the pest, each having constantly-changing variate-values.

2. Subtle interrelations of varying types occur among factors, and between the pest itself and its regulating factors. One example will suffice to illustrate the sort of interactions that can occur. It now seems reasonably certain that the effect of insecticides is governed by temperature (Pradhan and Srivastata, 1956; Pruthi,

1956a, b). However, it is also known that temperature, host density and parasite density all interact to regulate parasite oviposition rate. (Burnett, 1954, 1956). Hence, temperature, insecticide concentration, host density and parasite density all interact to control survival of a host.

3. Many factors regulating insect populations operate via more than one causal pathway. Temperature is typical of these. It can act directly on the pest, as a high or low lethal temperature. Also, in addition to its actions mentioned in 2 above, it can change development rate and therefore the time during which a pest is vulnerable to predators and parasites of a particular stadium.

Further, best control procedure from one short or long-term point of view may not be best from some other short- or long-term point of view. For example, spray treatment dosages must be carefully regulated so the chemical does not have deleterious side-effects on beneficial flora and fauna, and does not leave ever-accumulating residues that can poison man. Also, economic considerations place restrictions on the ultimate solution of the control problem.

It is not surprising that control of insects of any other pests is at present almost entirely an ad hoc procedure. Probably the efficiency, expense and permanency of currently employed control practices are all capable of considerable improvement. This improvement could be achieved if pest control became more scientific (i.e. practised according to principles embodying insight into the dynamics of pest populations).

Two kinds of problems confront us if we wish to employ mathematical models as a tool in economic entomology.

The first type might be referred to as general problems. For example, how, in principle does one go about developing a model from a body of data? What is the most efficient way to organize field sampling with a view to providing data from which to construct a model? What use are mathematical models, and how, in principle can we use them to attain various practical ends?

Secondly, we are confronted with a host of specific problems whenever we attempt to work out a model for a particular organism. These specific problems will relate to details of strategy to be used in sampling the organism, and particulars of mathematical procedures by which we express in equation form the effects of independent variables on dependent variables. For example, how should we express the effect of density on fecundity, of temperature and humidity on freezing, activity and development rate, and parasite densities on numbers parasitized? There is a growing literature on such specific problems and proposed solutions. However, the field is rapidly becoming more sophisticated, and it seems reasonable to suppose that all, or almost all the equations currently in use for handling specific phenomena will soon be replaced by more insightful versions. For this reason, any discussion of general problems will rapidly become rather dated if current specific equations (herein called sub-sub-models) for particular phenomena are woven too tightly into the fabric of the central argument.

This manuscript is intended only as a discussion of the aforementioned general problems. Specific forms of equations for particular phenomena have been used as sparingly as possible, and are intended only as illustrations of general points.

It is apparent by now, that in the next decade the whole array of field biological sciences on which economic entomology feeds will evince a much greater application of mathematics. This is exceedingly heartening, because the history of physics teaches us that a judicious blend of deduction and induction constitutes the scientific procedure most conducive to a rapid evolution of our

insight into nature. Theoretical work gives direction to the field work; empirical studies make the assumptions of the theoretical work more nearly realistic. Either without the other can become sterile, and can lead to painfully slow progress characterized by much time wasted following up blind alleys.

The primary aim in presenting the proposals in this paper has been to clarify the logical relationships between the data which the field biologist collects and the uses to which these data can be put, and to spell out the procedures by which the data can be put to these uses. A secondary aim is to familiarize economic entomologists with the literature of mathematics and statistics which is most useful for analysis of a typical pest problem.

The reader may wonder at this point if models of sufficient complexity to describe the subtleties of the field situation can ever be created and handled economically.

However, models of considerable sophistication have already been developed to describe marine fisheries (Beverton and Holt, 1957), growth of an organism (Weiss and Kavanau, 1957), the growth of bacterial populations (Moser, 1958), epidemic behavior (Bailey, 1957), evolution of a population (Wright, 1937), and physiological and biochemical processes (Johnson et al, 1954). Many of these models have been demonstrated to be realistic. (The aforementioned six publications collectively provide an excellent introduction to the spectrum of mathematical methods that have been employed to develop and solve models for biological phenomena).

Furthermore, highly complex problems of all types are becoming increasingly susceptible to solution with the growing availability of high-speed stored-program electronic computers (e.g. Box and Coutie, 1956) and extensions of existing branches of mathematics and statistics. In particular, since 1951 there has been very rapid development of a branch of mathematics invented specifically to deal with problems analogous to those in economic entomology. I refer to mathematical programming (Koopmans, 1951; Vajda, 1956; Bellman, 1957; Vazsonyi, 1958; Dorfman, Samuelson and Solow, 1958).

The development of entomology in general and applied entomological research teams in particular is characterized by a tendency to fragmentize. What started as an investigation by a single field ecologist often evolved into a team project with ecologists, physiologists, biochemists, toxicologists, nutritionists, statisticians and other specialists participating. However, typically, no one on such teams is responsible for fitting all the resulting research findings together in a form which produces a model useful for yielding precise quantitative insight into optimum control. It would appear that the time for such synthesis is long overdue.

2. THE NEED FOR MATHEMATICAL MODELS OF PEST POPULATION DYNAMICS

Developing and processing suitably realistic mathematical models of insect pest population dynamics is a time-consuming and expensive business, particularly when the cost of electronic computation is considered. Hence it behooves this author to explain in unambiguous terms how the payoff from such activity justifies the attendant expenditure of effort and money.

The first use of a mathematical model is that it can show how to get best possible control, for any set of weather conditions, using control agents and techniques already available. If survival of a pest was governed by only one factor, we could make a graph of "proportion surviving" against values of the factor, and

find the value which minimizes survival, by inspection and interpolation. However, typically survival is under the influence of several factors, and the graphical method does not suffice. Hence mathematical methods of determining how to minimize survival must be used. Such methods are outlined in chapter 6. The more complex the control problem, the more likely we are to need mathematical tools to provide objective answers as to the best control procedure. With evidence of insecticide resistance constantly accumulating (Brown, 1958), it becomes clear that future control programs are likely to be quite sophisticated. We can expect to see balanced control using diseases plus insecticides, entomophagous insects plus insecticides and all three together, perhaps in conjunction with a cultural control program. It will be difficult to make intelligent decisions about the best procedure to follow, using only ad hoc decision-making rules.

Mathematical models can aid in the search for still better control agents than those currently available. In the first place, a model can be used in conjunction with the analysis of variance to determine which factors are indeed regulating the pest population, and hence are most worthy of further study. Secondly, a model can help us to see which particular attributes of a factor make the factor important, and determine the degree of importance as a regulating agent. In any instance we would procede as follows. First, we would arrive at a particular mechanistic model to express the effect of a given factor or factors (e.g. Watt, 1959a, 1960). Second, we would use statistical procedures such as those outlined by Watt, 1959a, to obtain appropriate parameter values for an array of cases. Third, we would try to relate the parameter values to some underlying measurable biological variable. Now knowing the underlying variable that regulates effectiveness of a factor, we can begin a systematic search for a most effective version of a factor. For example, if we discover that a constant measuring searching efficiency in an attack equation is indeed a measure of the relative size of two organs in the entomophagous species, we merely hunt for that species which according to the equation has the optimum relative size. A case in point concerns the toxicity of DDT-analogues. If we do not understand why one analogue is more toxic than another, we must find the best analogue by trial and error. This is not science, which procedes via a rational alternation of guessing and testing. If, on the other hand, we find out the relationship between molecular configuration and toxicity, as Riemschneider (1954) has done, then we can easily extrapolate to the characteristics of the most toxic analogue.

A mathematical model can clarify research objectives, by showing what variables need to be measured, and how, when and where they should be measured. For example, wind and rain can both drive insects off plants, either killing them in the process or starving them to death by taking them too far from their food for them to get back. However, there are quite different techniques for measuring, and expressing the effect of, each of wind and rain. Wind can be measured as miles of wind per hour, or maximum gust velocity within an hour. Rain can be measured as inches per 24 hours, or maximum drop velocity within a 24 hour period. Also, we can work with the measurement for the whole 24 hour period, or just a segment of the period, such as the daylight hours, or the hours when it was hot enough to maintain active feeding. Which measure of such variables is best may be determined very quickly using a model. The best measure is that which produces the highest partial correlation coefficient. If some important, underlying but unmeasured variable is having an effect, we can see clues to this fact in significant F-ratios for intermediate causes under regulation of this primary cause. In short, work with models can be a valuable, objective tool for guiding basic research into the most fruitful channels.

Finally, the history of physics has taught us that one of the most important consequences of model-building is that we may discover principles that are reasonable , but come as somewhat of a surprise. For example, this author, like many others, has wondered what it is in the environment of a pest that imposes the environment's carrying capacity for parasites. There are cases in nature where parasite population density seems to rise only to a certain level, as if it had encountered a ceiling of some sort. However, study of the equations in Watt, 1959a, shows that such ceilings need not be a function of the environment at all. In fact, the carrying capacity of an area for entomophagous insects can be due merely to intraspecific competition amongst such insects.

3. AN APPROACH TO THE PROBLEM OF DEVELOPING A MODEL

Before becoming engulfed in the mathematical, statistical and data-handling procedures by which a model is developed, let us chart an outline of our goal and the procedure by which we shall attain it. Initially, we will use a one-generation-per-year insect as illustrative material.

SYMBOLS

N_t	the density (number per unit area) of adult pests present immediately prior to oviposition in year t (it is assumed throughout that we mean one particular pest species).
N_{t+1}	the density of adult pests present at the corresponding time in year $t + 1$.
$T_{t,t+1}$	the trend index of the pest from the time N_t is assessed to the time N_{t+1} is estimated.
X_i	the variate value for the i th factor regulating numbers of the pest during the interval t to $t + 1$.
P_t	the proportion of N_t that are females and oviposit in t.
F_t	the mean fecundity in t.
S_E	proportion of eggs surviving to eclosion.
S_I . . . . . S_{VI}	survival of first . . . . . sixth instar larvae.
S_P	survival of pupae.
S_A	survival of adults up to and including the time adults oviposit in $t + 1$.

We seek an equation

$$T_{t,t+1} = \frac{N_{t+1}}{N_t} = f(X_1, X_2, ., X_i, ., X_n) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Only some of the n independent variables are subject to control by man, and others, such as those which describe weather, are not. By placing commas between the X_i's we indicate that we are not making any a priori assumptions about the pattern of interaction; we expect that it will not be possible to split the function into a sum or product of several functions, each independent of the others. Equation (1) may have side conditions (constraints) which are linear or non-linear, equational or inequational (of form $X_i \gtreqless a$), or it may describe allocation of control agents at successive points in time. The form of the side conditions will determine the mode of solution, and this matter will be discussed in chapter 6.

The problem of economic entomology, stated in terms of (1), is as follows. Given an array of variate-values for the subset of the X_i's uncontrollable by man, (e.g. weather factors), what values should be selected for the remaining X_i's over

which he does have control (e.g. stand factors, entomophagous insects and chemicals)? Equation (1) may be conceived as the locus of a point in n-space. In any given year, the positions which the trend index can take in this n-dimensional hypersurface are restricted by weather, and possibly certain other factors (e.g. stand factors in a forest). Given such restriction, what can man do to minimize the trend index? To answer this question for any pest, we must proceed to develop the form of (1), as completely as we can.

The most immediately obvious point to a biologist is that equation (1) needs to be carefully structured in phenological time, as opposed to chronological time. This is because the effect which a factor has on a population depends on when that factor operates relative to the maturation of individuals in the population. For example, heavy wind gusts are invariably important if they occur at a certain stage in development of some forest insect larvae (i.e. when tiny caterpillars are migrating, by dangling at the end of a silk thread). Knowing only that the gusts occurred in the last two weeks of June does not supply the necessary information. Similarly, egg predators and parasites will no longer be important to the pest species when it hatches. Young fruit insect larvae are safe from spiders once inside the fruit.

Hence we need to structure (1) as in (2)

$$T_{t:t+1} = P_t F_t S_E S_I \ldots\ldots S_{VI} S_P S_A \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Hereinafter (2) will be referred to as the model. Each letter of (2) may be equated to a sub-model, as in (3).

$$S_E = f_E ({}_E X_1, {}_E X_2, \ldots\ldots {}_E X_n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

In (3), f_E stands for the survival function for eggs, and ${}_E X_i$ denotes the variate value for the i th independent variable which can cause mortality when it operates during the egg stage.

Finally, equations such as (3) can be further broken down into sub-sub-models which are structural terms describing the operation of individual regulating factors during a stage in the pest's life cycle. These terms may include more than one variable, since, for example, temperature, humidity, parasite density and host density all regulate the numbers attacked by each parasite species.

We now proceed to the problem of obtaining the structural form and parameter values for sub-sub-models, and fitting these together in a biologically appropriate fashion to form the sub-models.

Preliminary testing to get a rough idea as to the form of the sub-model and sub-sub-models involves a great deal of sorting, as will be described hereinafter. Hence processing is facilitated at the outset if the data are transferred to decks of cards. If up to 1,000 units of data are to be sorted, hand-sorting cards may be adequate. For 1,000-3,000 cards, a needle-sorting card system may be most efficient, and for over 3,000 units of data, electro-mechanically sorted punched cards may be best. These figures are very rough, and will be modified by the number of items in a unit of data, and the complexity of the subsequent calculations to be performed on each unit. (Two sizes of printed hand-sort cards, with spaces for 18 and 34 variables, are available to personnel of the Canada Department of Agriculture from the author). In the case of punched cards, enough reference data should be punched on to each card to identify it completely if it should become mixed with some other deck of cards. Such mixing can occur very easily in a room where millions of cards are stored and handled.

We shall illustrate our remarks on the coding of relevant data onto cards by referring to a simple card (Fig. 1). This card does not correspond to any real

PEST DENSITY AT t_1: N_1	SURVIVAL $= \frac{N_1}{N_0}$	DENSITY OF PREDATOR #1 AT t_0	DENSITY OF PREDATOR #2 AT t_0	DENSITY OF PREDATOR #3 AT t_0	DENSITY OF PARASITE #1 AT t_0	WIND GUST INDEX
PEST DENSITY AT t_0: N_0	ORCHARD		TREATMENT (SPRAY SCHEDULE)			HOURS CLOUDY, t_0 TO t_1
THIS SAMPLING DATE: t_1	PLOT			YEAR		INCHES OF RAIN, t_0 TO t_1
LAST SAMPLING DATE: t_0	TOPOGRAPHY	ISOLATION OF ORCHARD	DISPERSION OF FOOD PLANTS	FOOD QUALITY INDEX	TEMPERATURE INDEX	MEAN RELATIVE HUMIDITY, t_0 TO t_1

Fig. 1. Sample code used to enter data onto hand sort card.

situation of which the author is aware, but contains a variety of variables deliberately selected to illustrate various methods of coding data. One deck of cards will be required for each stadium of the pest.

Population estimates of the pest and its entomophagous controls should be obtained using sampling schemes as explained by Morris (1955). With respect to predators, different ages of predators (e.g. nymphal and adult Hemiptera) have age-specific consumption rates, and should be recorded and coded on the cards separately.

For parasites, the measurements of the pertinent variables required in the following analysis are not the effects of the variables, but rather the appropriate values for the variables themselves. That is, if the mortality due to *Glypta fumiferanae* was .20 and the density of this parasite was 146,000 per acre, figures of the latter type are the ones that must be entered into the equation. The reason for this is that the .20 is not only a function of (determined by) the parasite population density, but also the host density, and possibly other factors such as the total physiologically relevant heat during the searching period. Hence, it is relatively useless as an index of year-to-year changes in parasite abundance, and should not be used unless the density figure is unavailable.

It sometimes happens, however, that the appropriate measure is simply not obtainable. For example, a species may be highly secretive in habit, or may inhabit niches inaccessible to samplers, or there may be other technical difficulties, as in studying a virus epidemic. There are two sorts of indirect methods for assessing the mortality factor in such cases. In the case of parasites, for example, if a parasite has only one host, or is largely confined to one host, a rough measure of the parasites available to attack this host generation is the density of the last host generation multiplied by its percentage parasitism due to the species in question. A second method, which might be called "short-circuiting" is useful in handling disease statistically. Suppose we diagram a possible causal system (Fig. 2).

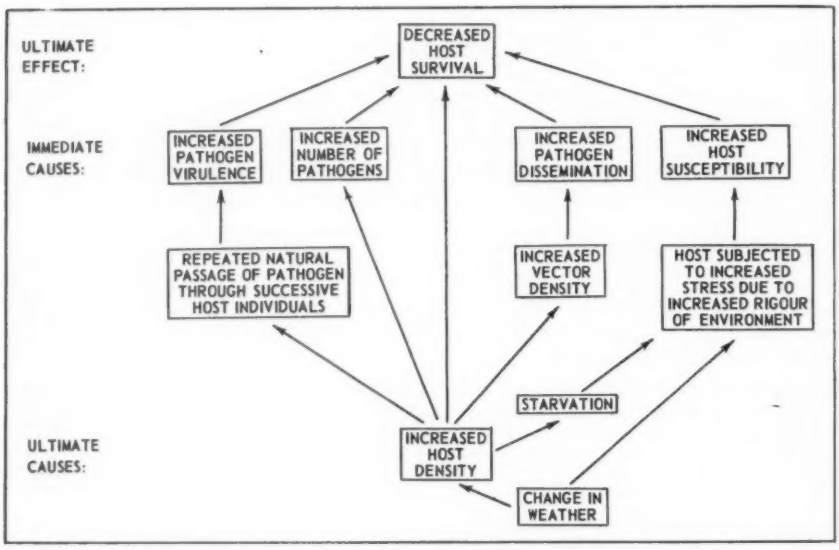

Fig. 2. Possible system of causal pathways by which effect of disease on survival is regulated.

While the four immediate causes which may determine effectiveness of a pathogen may not be readily measurable, they, in turn, are the effects of other causes, say weather and host density. Hence we may be able to omit disease factors from our equations, and express host survival as a function of the factors which affect it indirectly by affecting disease. Indeed, it is difficult to know a priori precisely how to diagram a causal system for an epizootic wave, or which variables to code, since insight into insect epidemiology is in an embryonic state (Steinhaus, 1954).

Wind gust index*, dispersion of food plants, isolation or orchard and topography are all coded, since some combination of these will determine mortality due to being blown off the host plant. In connection with the coding of such variables, two comments seem relevant. First, a scheme can be devised for quantifying any qualitative variable, such as topography. Such schemes will be illustrated in forthcoming publications of the Green River spruce budworm group (Morris and Miller, 1954). Second, where a variable has been measured as a number, this number should be entered on the card. To use a numerical code in place of a number is to make inefficient use of available information.

Table I illustrates inefficient use of data. In this case, the raw population densities should be entered on cards. Which data, and how many variables are coded must depend on the knowledge field biologists have of their pest's ecology. For example, if parasite 1 on the card in Fig. 1 searched for and attacked 2nd instar hosts, but parasite larvae were not fatal until the 6th host instar, then the relevant independent variable is the parasite density which the host was exposed to during its 2nd instar.

One further point needs to be made in connection with entering the data on cards. The time interval between sampling may be long, and predator densities can change rapidly through the interval. In such cases, rather than developing

*Most anemometers measure miles of wind per hour, but maximum gust velocity, probably a more biologically-relevant variable, is measured by a Dines anemometer, an Aerovane (Bendix-Friez Instrument Co., Baltimore 4, Md.), or a Beckman Whitley anemometer.

TABLE I

Example of an Incorrect Method for Coding Data onto Sorting Cards.

Code Entered on Card	Legend: Population Density per 10 sq. ft. of Foliage of Predator Species No. 4
0	0 – 24
1	25 – 74
2	75 – 174
3	175 +

the analysis using predator counts at the beginning of the period, the average density during the interval may be more appropriate. An estimate of the figure to be coded may be obtained as follows. If P_0 denote predator density at the beginning of the period, b the decay coefficient in $P_t = P_0 e^{-bt}$ and t_1 the length of the interval in days, we may write, for the average density during the interval,

$$\frac{\int_0^{t_1} P_0 e^{bt}\, dt}{t_1} = \frac{P_0}{bt_1}\left[e^{bt_1} - 1\right]$$

The expression $[\ln(P_t/P_0)]/t = b$ can be used to obtain b.

When decks of cards have been prepared for all stadia requiring separate analysis, we can proceed to use them for sorting. In order to present a lucid explanation of the technique and aim of sorting and subsequent operations, it is necessary first to discuss the structure of sub-models.

Let us consider the proportion of III-instar larvae, S_{III}, surviving to the fourth instar, where N_{III} represents the number of larvae present at the beginning of the third instar, and M_i is the number that would die due to mortality factor i if that were the only factor operating. We shall begin by considering the case in which only two mortality factors are involved.

One logical possibility is that the two agents operate independently, and are synchronous. For example, if a larva can be killed by weather or spiders, and the action of neither agent increases or decreases the probability of action by the other, we invoke the multiplication theorem of the calculus of probabilities to express the net effect. This states that the probability of neither of two events occurring is equal to the product of the two probabilities of non-occurrence. Hence,

$$S_{III} = \left(1 - \frac{M_1}{N_{III}}\right)\left(1 - \frac{M_2}{N_{III}}\right) \quad \ldots\ldots (4)$$

A second conceivable eventuality is that two mortality factors operate mutually exclusively. For example, parasite females of some species almost invariably avoid ovipositing in a host already containing an egg or larvae of another parasite species. Here

$$S_{III} = 1 - \left(\frac{M_1 + M_2}{N_{III}}\right) \quad \ldots\ldots (5)$$

A third imaginable contingency is that a second mortality factor makes no additional contribution to the mortality caused by the first. An instance of this

is provided by parasites which attack a host only after it has been committed to certain death by another factor, such as disease. *Kellymya kellyi* (Ald.) may be such a parasite. Smith (1958) states, "The almost complete absence of *K. kellyi* in collections of living grasshoppers may be due either to the rapidity with which infested grasshoppers are immobilized or to the fact that the parasite is not particularly effective except in grasshopper outbreaks, when, because of its rapid larval development and semisaprophagous habit, it is able to take advantage of the abundance of molting, weakened, and dying hosts".
In this case,

$$S_{III} = 1 - \frac{M_1}{N_{III}}$$

A fourth possibility is that the probability M_2/N_{III} is a function of the probability M_1/N_{III}. For example, caterpillars that had been subjected to prolonged stress from cold weather which had killed M_1 of them directly might suffer further apparent mortality M_2 due to stress-induced lowered resistance to disease. Here we may write

$$S_{III} = \left(1 - \frac{M_1}{N_{III}}\right)\left(1 - f\left\{\frac{M_1}{N_{III}}\right\}\right) \quad \ldots\ldots(6)$$

Another class of cases in which two mortality factors combine their effects occurs when the two factors are not synchronized. The simplest case of non-synchrony, that in which a second factor follows after the action of a first is complete, was first considered by Thompson (1928). Where M_2 follows M_1, he noted (converting his equations and symbols into the form we have been using) that

$$S_{III} = 1 - \frac{M_1}{N_{III}} - \left(1 - \frac{M_1}{N_{III}}\right)\frac{M_2}{N_{III}}, \quad \ldots\ldots(7)$$

which is equivalent to (4).

Any number of other interactions may be conceived, in which various combinations of the aforementioned cases occur. Also, in nature there are cases in which two mortality factors overlap, in time, so that the second begins to operate before the first is completed. The mathematical mode of handling such cases will depend on the form of their mortality functions.

An obvious generalization from (4) and (7) to n mortality factors is (8).

$$S_{III} = \left(1 - \frac{M_1}{N_{III}}\right)\left(1 - \frac{M_2}{N_{III}}\right) \ldots\ldots \left(1 - \frac{M_n}{N_{III}}\right) \quad \ldots\ldots(8)$$

Thompson (1928) distinguished "apparent" and "real" mortality due to a destructive factor operating on a population. These two terms, used frequently in this paper, can be defined in terms of (8).

"Apparent" mortality due to factor 1 is the proportion M_1/N_{III}.

"Real" mortality due to factor 1 is the proportion

$$\frac{M_1}{N_{III}}\left(1 - \frac{M_2}{N_{III}}\right)\left(1 - \frac{M_3}{N_{III}}\right) \ldots\ldots \left(1 - \frac{M_n}{N_{III}}\right).$$

Morris (1957) has recently discussed the interpretation of mortality data.

Any knowledge based upon field or laboratory experience will be useful in structuring the sub-model (entering the sub-sub-models correctly relative to each other). The biologist should be alert to certain possibilities he may encounter.

One noteworthy feature of sub-models for survival is that the same factor may operate in a causal system in two different ways, and hence might have to be entered in a sub-model, or even a sub-sub-model as parts of both a divisor, and an exponent, say. A simple example is the equation describing the effect of host density, N_o, parasite density, P, and number of eggs per female parasite, K, on N_A, the number of hosts attacked. Where a and b are constants,

$$N_A = PK\,(1-e^{-aN_oP^{1-b}}), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

(Watt, 1959a). Here N_A increases with P, because the more parasites there are, the higher the collective attack rate must be, but at the same time, increasing inter-attacker competition diminishes the attack rate, giving rise to the term $\exp - (aP^{-b})$.

Increasing host density, N_o, also has opposed effects. As N_o becomes larger, hosts are easier to find per attacker, and N_A increases. However, a given population of parasites has a finite complement of eggs to oviposit, and once host density passes over a certain level (PK) relative to parasite density, there will not be enough eggs to go around the hosts. In such cases N_A will level off at the asymptote PK. Here there is "passive group competition" amongst the hosts in the Schmalhausen (1949) sense, but it has a positive and negative component.

Rather than citing additional instances of single variables exhibiting multi-channel operation in systems of causal pathways, the reader is invited to reflect on the matter for himself, using Fig. 2 and equation (9) as heuristic devices. It will become clear that indeed almost all types of environmental factors can enter sub-model equations in two or more ways.

Three different strategies may be employed to particularize the form of sub-models, and obtain values for their parameters. In this book they will be referred to as inductive, deductive, and mixed approaches. While the second of these represents the goal of work on natural pest populations, the third, an intermediate step towards the end objective, should be explained in terms of the steps which precede and follow it, since it contains features of both. Hence it is explained last.

If the entomologist begins structuring a sub-model with no insight into the mechanics of the causal pathways he wishes to describe, he is committed to an inductive equation as a first step. Books by Ezekiel (1941), Deming (1938), Quenouille (1952) are particularly helpful in summarizing information on procedures for formulating, developing and testing such equations. An example of such an equation is

$$S_{III} = a + bX_1 + cX_2 + dX_3 + eX_4, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

where the X_i's can represent a single factor, a factor squared or cubed, or a product of two factors, etc.

If, as would ideally be the case, an entomologist has complete understanding of the web of causal pathways regulating S_{III}, he can write a deductively-based equation, using assumptions tested inductively, which accounts for all the year-to-year and place-to-place within-a-year variation in S_{III} not accounted for by chance. Not enough is known yet to write an example with any degree of assurance that the equation has a counterpart in reality. (The whole aim of this monograph is to interest workers in remedying the deficiency). However, enough equations albeit still of a relatively unsophisticated type, have been developed and tested to indicate the nature of such structural forms.

Assuming,

(1) that the only causes of mortality affecting S_{III} are parasitism, predation and intraspecies competition, and

(2) that the appropriate equation corresponding to $S_{III} = \left(1-\frac{M_1}{N_{III}}\right)\left(1-\frac{M_2}{N_{III}}\right)$ where M_1 and M_2 are apparent mortalities caused by parasites and predators is

$$S_{III} = \left[\frac{1-P_1K_1\left(1-e^{-a_1N_{III}P_1{}^{1-b1}}\right)}{N_{III}}\right]\left[\frac{1-P_2K_2\left(1-e^{-a_2N_{III}P_2{}^{1-b2}}\right)}{N_{III}}\right]$$

(3) the appropriate equation for mortality caused by intraspecies competition is

$$S_{III} = 1-\frac{N_{III}e^{-c}}{N_{III}} = 1-e^{-c},$$

after various authors, from Baranov (1918) to Beverton and Holt (1957),

(4) that K_1, a_1, b_1, K_2, a_2, b_2 are all similar functions of temperature, the form of the function at the present stage of insect physiology being best given, e.g., by

$$K_1 = K_0e^{-dx^2},$$

Pradhan (1945, 1946a, b), Matsazawa, Okamoto and Miyamoto (1957), where K_0 is the maximum K at any temperature, d is a constant, and x the difference between the temperatures producing K_0 and K_1, then we can write

$$S_{III} = \left[\frac{1-P_1K_{10}e^{-d_1x_1^2}\left(1-e^{-a_{10}e^{-d_2x_2^2}N_{III}P_1{}^{1-b_{10}e^{-d_2x_2^2}}}\right)}{N_{III}}\right]$$

$$\left[\frac{1-P_2K_{20}e^{-d_4x_4^2}\left(1-e^{-a_{20}e^{-d_5x_5^2}N_{III}P_1{}^{1-b_{20}e^{-d_6x_6^2}}}\right)}{N_{III}}\right]$$

$$\left[1-e^{-c}\right] \qquad (11)$$

Finally, if an entomologist had adequate insight into a field population of pests to write some terms such as those in (11), but could not do this for all factors, his final equation would contain some terms as in (10), and some as in (11).

Sufficient background material has now been given so that we can proceed to an explanation of how we would use data, such as that in Fig. 1, on cards to develop any of the three types of models.

4. DEVELOPMENT OF AN INDUCTIVE MODEL

The first step in developing a model of the first, or purely inductive type, is to sort the cards by each of the independent variables in turn. A convenient way of doing this is to rank the cards by "Density of Predator No. 1 at t_0", say (as in Fig. 1), then divide the ranked deck into ten equal sub-decks. For each sub-deck the mean survival and mean predator density is tabulated. Then mean survival is plotted against predator density. Such tables and plots are made for each independent variable.

TABLE II

The Effect of Pest Density at t_0, Predator No. 3 density at t_0, and Wind Gust Velocity on Pest Survival During Third Larval Instar.

	Survival of III-Instar Pest Larvae		Pest Density at t_0		Predator No. 3 Density at t_0		Wind Gust Velocity Index		Rank of Specified Division		
Number of cards in sub-sub-Deck	Total For 5 Cards	Mean For 5 Cards	Total For 5 Cards	Mean For 5 Cards	Total For 5 Cards	Mean For 5 Cards	Total For 5 Cards	Mean For 5 Cards	Sub-Deck	Sub-sub-Deck	Sub-sub-sub-Deck
5									5	5	4
5											3
5											2
5									5	4	1
5											4
5											3
5											2
5									5	3	1
5											4
5											3
5											2
5											1
5									5	2	4
5											3
5											2
5											1
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5				(missing values)					1	1	4
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
500									1		1

[1] Total no. of cards in deck = 5 cards per sub-sub-sub-deck
× 4 sub-sub-sub-decks per sub-sub-deck
× 5 sub-sub-decks per sub-deck
× 5 sub-decks per deck
= 500 cards.

There are two reasons for sorting to obtain mean values for plotting, rather than plotting the raw data. First, if there were 500 cards coded as in Fig. 1 in the deck, there would not be enough room for all the plotted points on any graph of reasonable size. Second, suppose we wish to determine the effect of "Wind gust index" on survival. Each value of survival measured from field data represents the effect of not only wind gusting, but also a number of other independent variables, and chance. Hence even if wind gust velocity was quite important in determining survival, this effect could be masked in the raw survival ratios. By pooling survival and wind gust data from 50 cards, we produce a mean survival ratio from which the influence of other independent variables has been eliminated, to a considerable extent. That is, for any given range of values of wind gust velocity, as represented in one of the sub-decks of 50 cards, there would be a probability distribution of values for, say, temperature. By pooling the 50 cards on the basis of wind, not temperature, we should arrive at an average figure for temperature that will be approximately the same for each of the 10 50-card sub-decks, assuming there is no strong correlation between temperature and wind gust velocity. Obviously we are discussing a method which gains in strength as the number of cards increases, and as the sampling error attached to the data on each card decreases.

After all independent variables have been plotted as discussed above, it should be clear which variables are the most important in accounting for year-to-year and place-to-place within-a-year variation in survival for a given stadium. The graphs yielding the highest correlation coefficients will be those for which the slopes of the lines are greatest, and the scatter about the lines is least.

The next step is to determine the pattern of interaction amongst the most important variables. The cards have now been sorted into sub-decks, as previously outlined, sorting having been done on the most important independent variable. Then each sub-deck is sorted into sub-sub-decks, on the second most important independent variable, using the same technique (ranking, then dividing the sub-deck into n equal sub-sub-decks). Finally, the sub-sub-decks are further subdivided, sorting on the third most important independent variable. Tabulating appropriate mean values for all the sub-sub-sub-decks will yield a table such as Table II.

There will now be 100 sets of values to plot. These can be plotted as five points for each of five curves in each of four graphs, as in Fig. 3. Examination of plots such as Fig. 3 will yield insight into the manner in which three independent variables interact to govern the dependent variable. The simplest case we can expect to find is that in which the net effect contains no interaction term. This is the situation depicted in Fig. 3. In this figure, the total effect of all independent variables is merely the sum of the effects of each. If, on the other hand, the net effect involves interaction terms, the corresponding plot will be as in Fig. 4. The equation corresponding to Fig. 3 is (12), and that corresponding to Fig. 4 is (13), where S_{III}, N_{III}, P and W denote survival of third instar larvae (to the time of moulting), number of larvae at the beginning of the third instar, density of predator No. 3 at the beginning of the past larvae's third instar, and wind gust velocity index, respectively. The small letters represent constants.

$$\ln S_{III} = a - b\,N_{III} - cP - dW \quad \ldots\ldots (12)$$

$$\ln S_{III} = a - b\,N_{III} - cP - dW + fN_{III}P \quad \ldots\ldots (13)$$

It should be noted that figures based on actual field data would never appear as in Figs. 3 and 4 when first plotted. Such smooth curves would result only

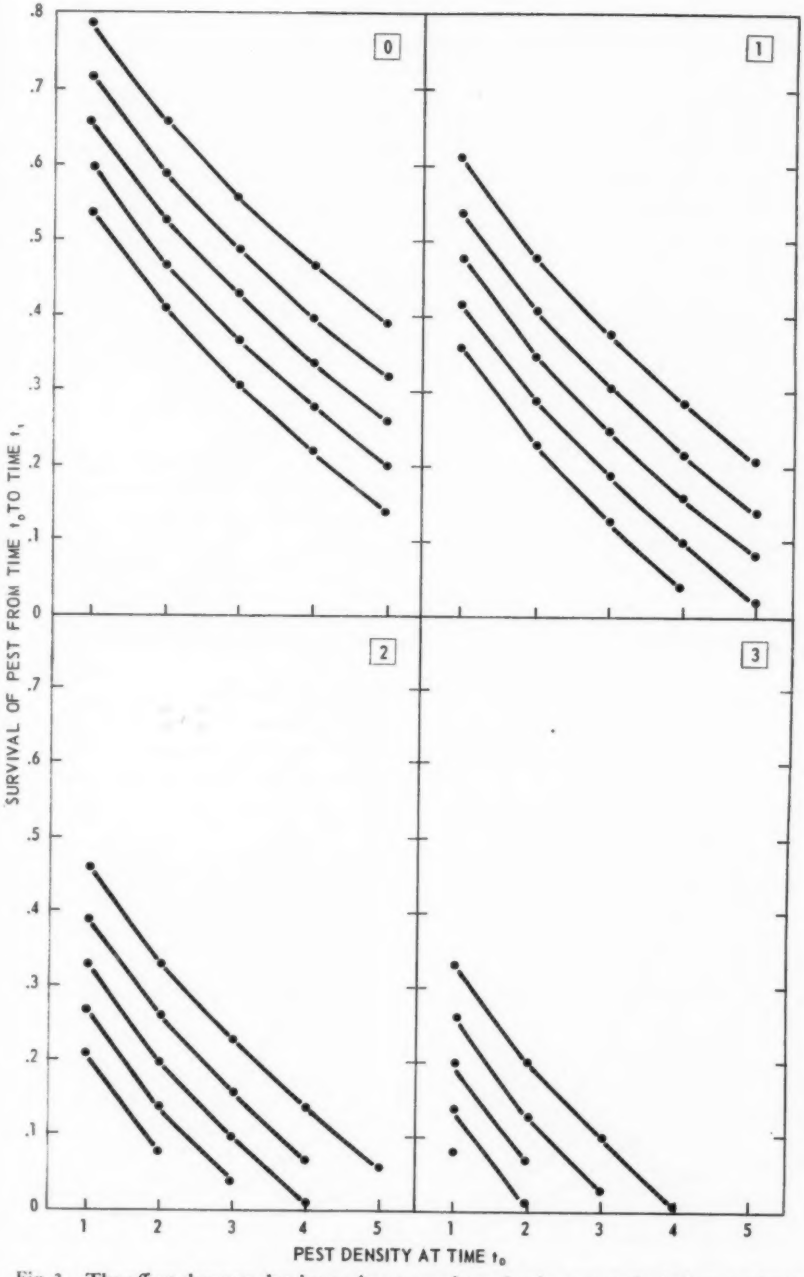

Fig. 3. The effect that pest density at time t_0, predator density at t_0 and wind gust velocity would have on survival of a hypothetical pest population if the net effect of the three independent variables was equal to the sum of their effects. No random effects were included in model (equation 12 in text). The four sets of curves labelled 0, 1, 2 and 3 represent wind gust velocity indices of 0, 1, 2 and 3. Within each set, the uppermost of the five lines represents, one predator, the next two predators, and the bottom line five predators per unit area.

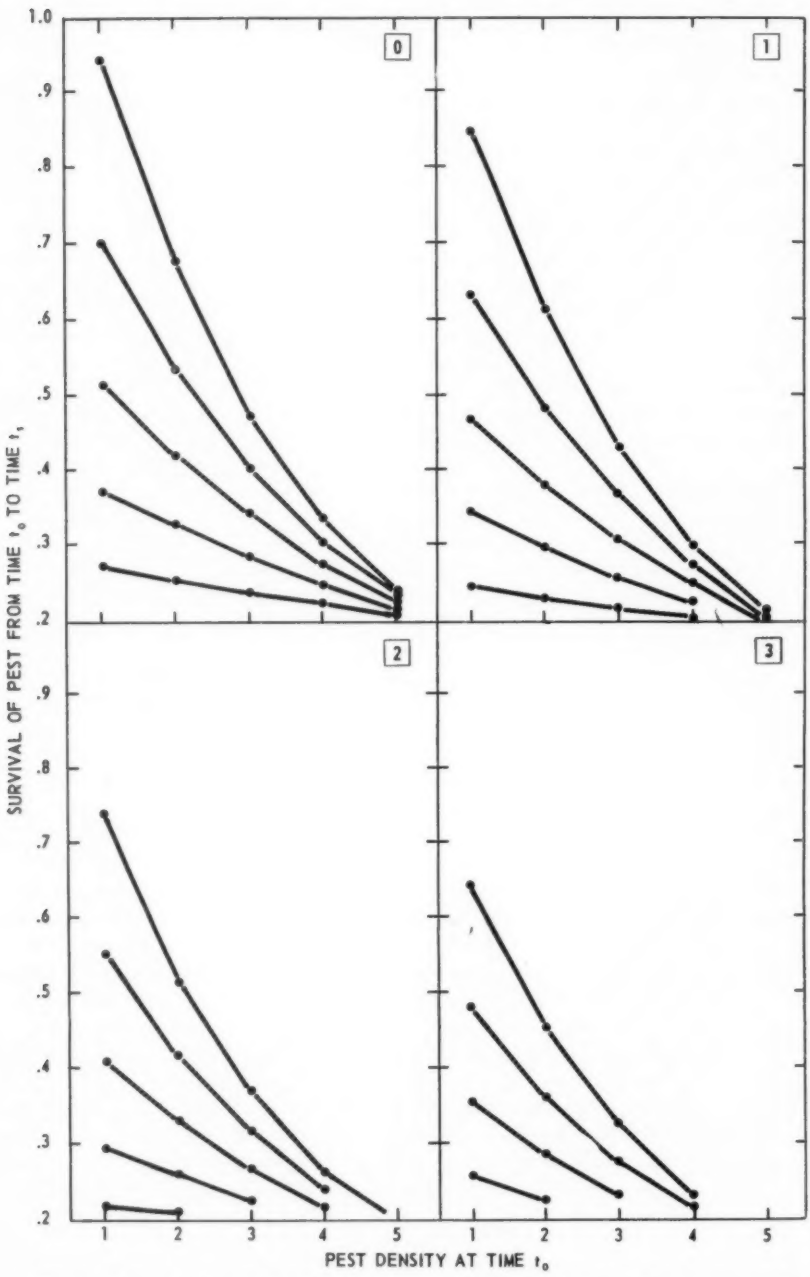

Fig. 4. Same situation as for Fig. 3 was assumed, except that equation 13 was used for model, with interaction between pest and predator densities.

from multidimensional graphic smoothing as outlined by Ezekiel (1942, chapter 21).

Once sets of curves such as those in Figs. 3 and 4 have been plotted, the next problem is that of determining from inspection of the curves, and subsequent tests, the nature of the equation depicted by the curves. This problem has two aspects. First, we must determine the appropriate transformations for the dependent and independent variables. (Should X_2 be incorporated into the equation as X_2, $\log X_2$, $\frac{1}{x_2}$, $\log (1/X_2)$, or in some other form?). Second, what is the pattern of interaction of the independent variables?

Considering Fig. 3, we note, first, that survival does not drop retilinearly with increasing pest density. However the curves do not bend in sharply towards the origin, so we are led to suspect a form

$$\ln S_{III} = a - bN_{III}, \quad \text{rather than}$$

$$\ln S_{III} = \ln a - b \ln N_{III}.$$

Next, we notice that for each of the four panels representing various wind gust velocity indices, different predator densities do not change the shape of the S_{III} on N_{III} relationship. Similarly, the increasing wind gust velocities drop the curves equal amounts for equal increases in wind gust velocity. Evidently we are dealing with an equation such as (12). This would be tested by plotting

$$\ln S_{III} + b N_{III} = a - cP - dW,$$

which would yield, for different P and W combinations, a family of parallel straight lines.

Discussion of Fig. 4 will yield insight into the logic by which we determine the equation for a more complex relationship.

First, we consider the panel for a wind gust velocity index of 0. As in Fig. 3, for any line we can write the equation

$$\ln S_{III} = a - bN_{III}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

However, the Y-intercept, a, is a function of P. (The distances between the Y-intercepts of the lines in Fig. 4 are not equivalent, but the reader should note that the relevant variable is $\ln S_{III}$, not S_{III}).

We may write

$$a = c - dP, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$$

but it is obviously also true that P affects the slope of (14), and the simplest relation we can write for this is

$$b = f - gP. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

Now comparing the four sets of graphs in Fig. 4, we note that while the intercepts shift down with increasing wind gust velocity index, the slopes are unchanged. Hence

$$a = c - dP - hW \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (17)$$

and inserting (17) and (16) into (14), we get

$$\ln S_{III} = c - dP - hW - N_{III}(f - gP), \quad \text{or}$$

$$\ln S_{III} = c - dP - hW - fN_{III} + g N_{III}P,$$

which is equivalent to (13), and in fact is the law according to which Fig. 4 was derived.

More complex equations can be developed using analogous reasoning to cover more complex phenomena with more variables.

The functions depicted in Figs. 3 and 4 are extremely simple relative to others that are needed to describe biological phenomena. However, a large array of mathematical devices is available for both determining the appropriate form of transformation and studying patterns of interaction. For convenience, a list of these is assembled below.

(1) Graph papers have been developed to test a great variety of functional forms. Furthermore, various techniques of sorting and pooling data can increase

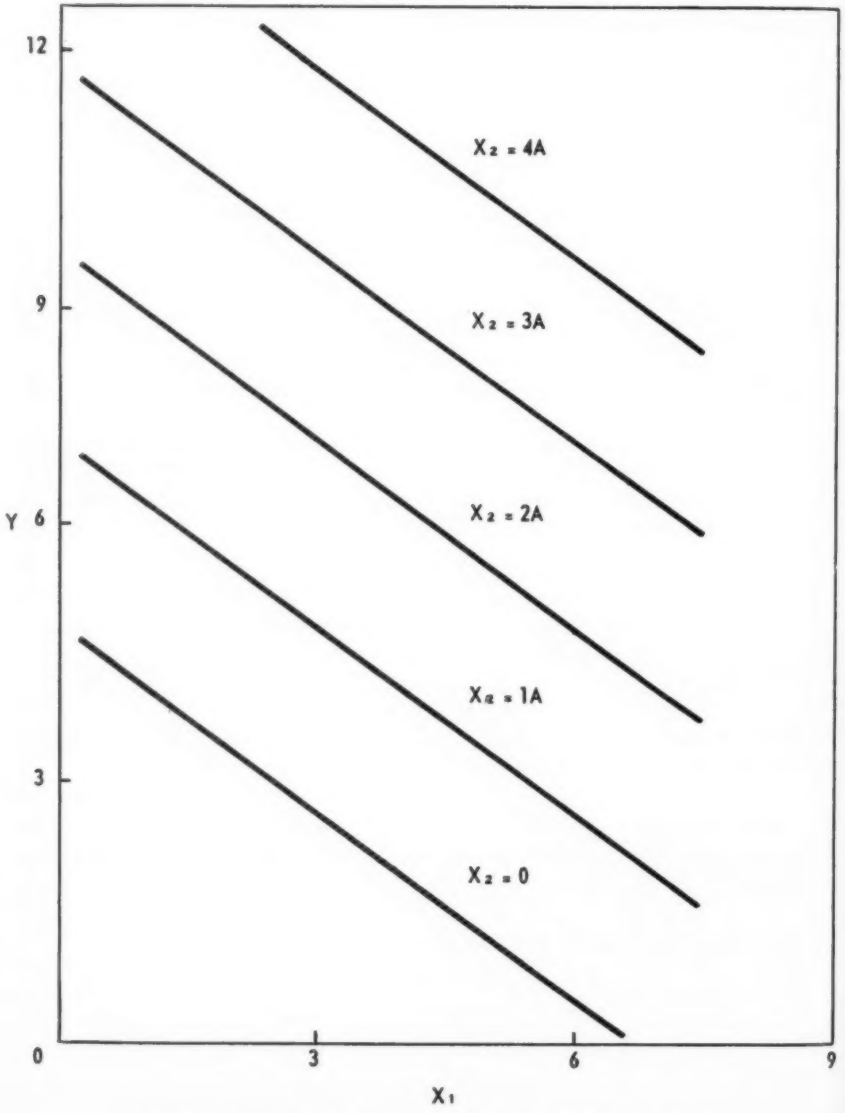

Fig. 5. Graphical test of hypothesis that $Y = a - bX_1 + cX_2$.

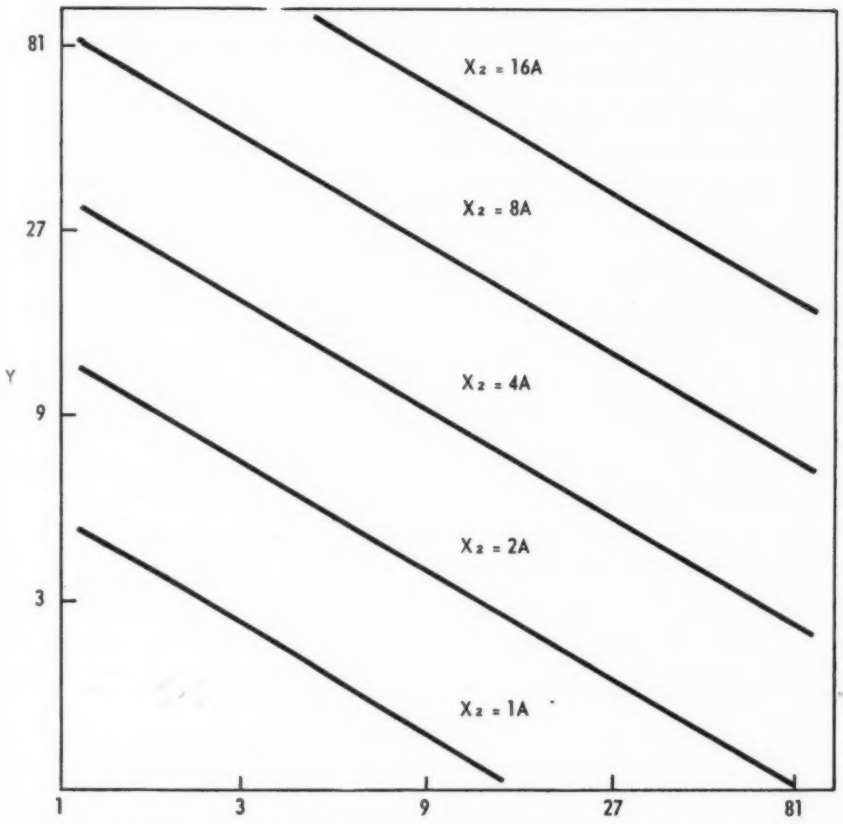

Fig. 6. Graphical test of hypothesis that $Y = AX_1 \cdot {}^aX_2{}^b$ or $\ln Y = \ln A - aX_1 + bX_2$.

the usefulness of papers already available. A variety of techniques can be developed to test 3-, 4- and 5-dimensional relations on a plane graph. Two simple examples, together with a reconsideration of Figs. 3 and 4, will serve as heuristic devices for the reader.

Suppose we are dealing with the relation

$$Y = a - bX_1 + cX_2. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

A simple means of determining if any given set of data conform to (18) is to sort the data into sub-sets on X_2, so that the mean values of the several piles are in arithmetic progression. Plotting the data for each sub-set against X_1 will then yield a series of parallel and equidistant straight lines, as in Fig. 5, if the data indeed conform to (18).

Now a next logical step is to use this technique for forms such as

$$Y = AX_1{}^{-a}\, X_2{}^b \quad , \text{ or}$$
$$\ln Y = \ln A - a \ln X_1 + b \ln X_2. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (19)$$

If data suspected of conforming to (19) are sorted into subsets on X_2, with X_2 means in geometric progression, the resulting plot will resemble Fig. 6 if (19) actually applies. The same principle can be extended to other types of trans-

formation, and additional numbers of variables. For example, plots of 4-variable relations can be made on one graph, using several sets of lines, one color for each value of the 3rd independent variable.

(2) Possibly the most widely-used technique for determining the correct form of a model is to transform the equation into straight line form. Then, when the equation is plotted, we know that the original (untransformed) form was applicable if the plot forms a straight line. This topic will be expanded in the next chapter, on deductive model-building. The most complete discussion of transformations this author has seen is by Running (1917), a rare book well worth the trouble involved in obtaining it. An excellent modern treatment on statistical aspects of transformations is by Hald (1952, pp. 558-570). Collections of techniques for testing hypotheses about the form of equations are given by Davis (1955), L. H. Johnson (1952), and Lipka (1918).

(3) A third method of determining which equation is appropriate to describe data is that of finite difference tables. This topic is discussed in many books, including Running (1917) and Worthing and Geffner (1943), but the basic idea will be given here.

Suppose we suspect that a body of data is described by $y = a + bx + cx^2$. Then, where $\triangle$ denotes an increment, and reasoning as in the calculus,

$$y + \Delta y = a + b(x + \Delta x) + c(x + \Delta x)^2$$

and after subtraction, and expansion

$$\Delta y = b\Delta x + c(\Delta x)^2 + 2c\,x\Delta x \quad \ldots\ldots(20)$$

The test is conducted by drawing up a table of values of $\triangle y$ and the differences between the $\triangle y$'s, for a constant increment $\triangle x$. Since $\triangle x$ will be kept constant, (20) may be rewritten as

$$\Delta y = a' + b'x \quad \ldots\ldots(21)$$

We can now repeat the procedure that gave rise to (21), and this will yield

$$\Delta^2 y = a'', \quad \ldots\ldots(22)$$

where $\triangle^2 y$ represents the differences between two successive $\triangle y$'s. This technique can obviously be extended to a great variety of types of functions. Many of the tests are presented in Running (1917).

(4) We can determine whether a term is an addend or a multiplicand in an equation as follows. Suppose we wish to determine which is appropriate,

$$Y = a + bX_1 + cX_2, \quad \text{or} \quad \ldots\ldots(23)$$

$$Y = (a + bX_1)(c + dX_2). \quad \ldots\ldots(24)$$

Equation (23) will yield a straight line if $Y - bx_1$ is plotted against X_2. The obvious extension of this principle is outlined by Quenouille (1952, p. 27). Equation (24) will yield a straight line if the quotient $Y/(a + bX_1)$ is plotted against X_2. Obvious analytic extensions of this device will be presented in the next chapter.

Once the form of a survival sub-model has been determined, the next problem is to get the best possible estimates for the parameter values. It may be necessary to expand a survival sub-model, and collect like terms in order to convert it into a form suitable for regression analysis. For example, where X_1 and X_2 represent two stand factors, we may have decided that

$$\ln S_{III} = 1 - (a + bX_1 - cX_1^2) - (d + fX_2 - gX_2^2) + (a + bX_1 - cX_1^2)(d + fX_2 - gX_2^2), \quad \ldots\ldots(25)$$

to use a very simple case. Multiplying the last two terms together, and collecting like terms we obtain (using new parameter notation)

$$\ln S_{III} = b_0 + b_1X_1 + b_2X_2 + b_3X_1^2 + b_4X_2^2 + b_5X_1X_2 + b_6X_1^2X_2 + b_7X_1X_2^2 + b_8X_1^2X_2^2.$$

We can greatly reduce the number of symbols to be used hereinafter by writing the above as

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_8X_8,$$

where Y represents ln S_{III}, X_3 represents $X_1^2X_2^2$, and so on.

The aim of multiple regression analysis is to seek the values of b_0, b_1, b_2, etc., which best describe the way in which the X_i's govern Y. The traditional approach to this subject has been to seek the values of b_0, b_i etc. which together minimize the sum of squares of the vertical distances between observed and calculated values of Y.

At this point, the reader may be interested in references to background reading on regression methods. Excellent introductions to regression analysis are by Ezekiel (1941) and Quenouille (1952). A number of considerations may render the simplest approach inadequate. If each set of values of Y and the X_i's entered into the analysis is based on a series of observations, and the different series have different sample sizes and variances, then weighting will be necessary. Anderson and Bancroft (1952, p. 182) discuss the computational procedure. If various constraints are placed on the solution, extensions of the simple methods will be necessary, and Deming (1943) gives these. To get a deep understanding of the least squares method, Lindley (1947) is recommended. The methods discussed in most texts do not yield the best true values of the b_i's, but rather the best values for use in predictions. This point is developed further by Bartlett (1949b), Lindley (1947), and the papers in their lists of references. A number of authors including Anderson and Bancroft (1952), Bennett and Franklin (1954), Ezekiel (1941), Fraser (1958), Goulden (1952, Hald (1952), Kramer 1957) and Snedecor (1956) present complete outlines of computational technique. Programs for performing multiple regression analysis on stored-program electronic computers are available from all large computer-manufacturing companies.

The volume of computation involved in matrix operations will sometimes be great enough to require electronic computation. Since multiple regression analysis is a very commonly-used technique in a great many fields, standard programs for this routine have been prepared for most intermediate and large size electronic computers.

Computers confer three advantages: they save money, make possible types of calculations that would not otherwise be undertaken, and make data and data analyses available very soon after demand for them arises.

It seems advisable here to point out a type of pitfall that should be avoided in dealing with computers. Since they can do multiple regression analysis so quickly and cheaply, there might be a tendency to attempt to fit a function of form.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + b_5X_5 + \ldots . \\ b_nX_n$$

without having any real reason for including some of the X_i's except that one was embarking hopefully on a "fishing trip". At best, this will yield the experimenter a rough idea as to which variables are accounting for a significant proportion of the time-to-time and place-to-place variation in Y. At worst, it might be merely

a snare and a delusion. It would yield no insight into the mechanics of the system, some important variables might be shown to be significant when they were not, (they might be highly correlated with some unincluded variable which was important), and interaction effects might be obscured or masked. The *order* in which we remove independent variables from the matrix will govern the proportion of the variability accounted for by variables yet to be removed.

In the foregoing, we are merely reiterating a well-known fact, that a high correlation coefficient does not necessarily imply causation. In addition to cogent admonitions in various text-books regarding the interpretation of correlation coefficients (e.g. Wallis and Roberts, 1956; Finney, 1953), Berkson's (1958) paper is a singularly wise discussion of the significance of correlations.

We may check the interpretation to be placed on correlations either experimentally, or by manipulating survey data, as follows. For example, suppose we find that through a period of time, a certain predator mite increases while a prey mite it is known to eat decreases. This does not prove a causal relation. Whether or not there is a causal relation can be tested as follows: analyze the relation between prey trend index and predator density at any point in time, from place to place. If there is no relation, we know the high correlation between the two variables from time to time at all places was due to the correlation of each of them with one or more other variables that changed through time. For example, possibly the physiological optimum temperature for the predators was lower than that for the prey, so the biotic potential of the prey was higher in the early part of the season, while that for the predator was higher in the early autumn. If there is a relation, direct proof of predation is still needed to verify the correlation "evidence".

Now that we have concluded discussion of inductive models derived through multiple regression analysis many disadvantages of such models will be clear.

Firstly, while inductively-derived equations relate causes to effects, they yield no insight into the mechanics of the causal pathway. Hence, such models are not useful in systematizing the search for optimum control measures, because they do not expose the specific criteria which will aid the search.

Secondly, a linear multiple regression equation will be a woefully bad predictor if the true mode of action of relevant variables is markedly non-linear. In point of fact, as we shall show in the next section, the relationships between most, if not all biological variables are indeed strikingly non-linear, on the scales on which they are ordinarily measured.

Thirdly, in the opinion of this author, excessive dependence on linear multiple regression as a tool for biological research can have an insidious effect on the development of knowledge in a biological subject-matter field. The researcher should bear in mind that this statistical tool was not developed because linear models describe the behavior of complex systems in the real world. Rather, a linear additive model is theoretically straightforward to work with, and leads to a simple, if tedious computational procedure. Trying to cram biological facts into a model that is best from a statistical point of view, but poorest from a biological point of view is a procedure that can only result in leading a segment of biology into a theoretical cul-de-sac.

The only reason for including a discussion of multiple regression methods in this paper is that they constitute a natural stepping-stone to the more realistic models discussed in chapter 5. That is, results of work with multiple regression should serve as guides in preliminary work on construction of an analytical model.

5. DEVELOPMENT OF MIXED DEDUCTIVE-INDUCTIVE MODELS

The development of a deductive model to describe a complex phenomenon is quite a different matter from computing a regression equation, or inductive model. In the latter case, we try to relate all the year-to-year and place-to-place-within-a-year mortality to an array of independent variables, without postulating modes of operation for causal pathways. In essence, we merely settle for the best equation of an arbitrarily chosen simple type that accounts for the variation amongst an array of points. This is a blind interpolatory procedure.

In the former instance, we try to write an equation relating a small number of factors to the apparent mortality clearly ascribable to each of these factors. If we use pure deduction, the procedure is as follows. First, we make a set of reasonable assumptions about the effect of various factors on survival, fecundity and sex ratio of the pest. Second, these assumptions are expressed in the form of a set of equations, generally differential or partial differential equations. Then the equations are integrated, combined and solved, and a set of fairly elaborate deductions can be made. One of the best-known sets of such deductions is that of Volterra (1926). However, the inherent weakness in such pure deduction is that no matter how correct the logic by which the conclusions are deduced, the conclusions are no more correct than the basic assumptions. Hence, as the history of physics has taught us, mathematical model building is of limited use unless based on inductively verified assumptions. The optimum method of building a sub-sub-model, therefore, is to first guess at the mechanism of a causal pathway, then test the guess, or hypothesis, then refine the hypothesis on the basis of the test results. In effect, we use induction to determine the best assumptions, and deduction to explore the implications of the assumptions, so I am calling the results, "deductive-inductive models". This cycle of hypothesis, test and refinement of hypothesis is repeated in the classic scientific manner until we can express a scientific law that is most suitable for each sub-sub-model. Then we repeat the procedure on a larger scale to fit the resulting set of sub-sub-models together in the best possible manner to form a sub-model. Success will have been achieved, when we have a sub-model which explains how the observed apparent mortalities combined to yield the corresponding observed total real mortality.

It is the purpose of this chapter to outline as systematized as possible an approach to the foregoing partly intuitive process of model construction. (Ghiselin, 1952; Hadamard, 1945; and Poincare, 1910 constitute useful background reading on the process of mathematical creation).

The deductive-inductive approach requires an additional category of data to the inductive at the outset (it will require still more at later stages of development). We require apparent mortalities for each major factor (parasite, predator, disease, weather or stand factor, etc.). The reason why we must work with apparent, as opposed to real mortalities at this stage in model construction, is that this is the only way we can relate a segment of mortality to the factor which caused it with any assurance that the magnitude of the effect is accurately measured.

We now digress briefly to outline methods of measuring apparent mortality, for various types of factors.

Apparent mortality due to parasites may be determined by dissection or rearing in the laboratory. A detailed discussion of the technical problems encountered is given by Miller (1955).

It may be difficult or impossible to measure apparent mortality caused by predators. One approach is to enclose sections of the natural habitat, thereby

excluding predators (Dowden et al, 1953; Smith and DeBach, 1942), but this may prove to be too time-consuming and expensive, and may be modifying the environment in other ways. An alternative may be to measure real mortality caused by predators, by relating deaths of prey, or decline in numbers of prey from one time to another, to predator densities, using regression methods. One would have to take care to design a sampling survey which would effectively separate deaths due to predation from deaths due to other causes.

Disease can probably best be studied by evaluating apparent, rather than real mortality due to this cause. Care has to be taken lest handling and crowded confinement of insects increases the possibility of disease.

Dispersal losses, caused by dispersal of insects to a site that has no food, or is otherwise unsuitable, may be found by subtracting the number found at a late stage from a number estimated at an earlier stage in the same place. Miller (1958) discusses losses due to dispersal and miscellaneous other causes.

In the case of most other types of mortality factors, such as freezing, stand factors, crowding, and application of chemicals, it may be difficult to obtain apparent mortality estimates. As in the case of predation, it will probably be necessary to relate a particular segment of total real mortality to the variate-value for each of these factors, using multiple regression methods. The fact that we are able to measure apparent mortalities for some factors, but must settle for real mortalities in the case of others, is what increases the importance of the mixed type models discussed later in this chapter.

We begin to develop deductive-inductive sub-sub-models by plotting the apparent mortality caused by each factor against values of that factor. This is done for each factor. In the case of the independent variables for which we can not measure apparent mortalities, we must try to build deductive models insofar as possible using total real mortality data. The cards used in sorting, from which tabulations and the immediately aforementioned plots are obtained, should not only contain reference data (plot, year, etc.), apparent mortality due to X_i, (the i th independent variable) and the value of X_i. They should also contain data on all other independent variables which can modify the action of X_i.

The experimenter now carefully examines the data he has plotted, and asks himself, "What type of law seems to be governing the relation between apparent (or total real) mortality, and X_i?"

There are two means of answering this question. One method is to determine the nature of the law at work by analyzing the graph, and the process. The second is to compare the graph, and the process, with graphs and processes for which equations have already been derived. We shall first consider how to derive equations for phenomena by analyzing processes, and then we shall consider a number of equations that have already been derived, and that will be useful in reasoning by analogy.

In point of fact, any experienced mathematician would be likely to derive a law to describe a simple phenomenon so quickly, that it would be clear he had attained his end by intuition, rather than reason. However, this intuition is at least in part, and in most cases, the product of experience, as opposed to inherited characteristics. This author is of the opinion that the process of mathematical model-building, while essentially a creative act, can be taught, if a sufficiently penetrating analysis is made of the process.

The first step in arriving at a model, after consideration of a process and graphs of the process, is to determine in which of several broad categories a suitable descriptive equation would be found. Then, narrowing our attention to the

appropriate category of equations, we can initiate a systematic logical procedure to particularize the form of the equation. There is nothing essentially novel in this process, since it is analogous to the routine by which one would find an integral in a large table of integrals. Such tables as those of Dwight (1947) are organized to permit a systematic search, which proceeds from a broad category to a specific question.

All biologists are familiar with keys used to determine the identity of a specimen in hand. We can use the following analogous key to determine which broad category of equations contains the appropriate equation or equations to describe the phenomenon under study.

1. Independent variable Y can be thought of, at least initially, as being a function of only one other variable. ORDINARY DIFFERENTIAL EQUATION 2
   Independent variable Y is obviously a function of two or more independent variables. PARTIAL DIFFERENTIAL EQUATION(S). It may be advantageous from the standpoint of logical analysis to consider a multivariable situation as being describable by a single ordinary differential equation at this stage. The integral equation ultimately arrived at can be obtained either way.
2. It is not necessary to use derivatives of a derivative (e.g. d^2y/dx^2) in basic equation. FIRST ORDER 3
   It is necessary to use derivatives of higher order than one (such as d^3y/dx^3) in basic equation. ORDINARY DIFFERENTIAL EQUATION OF ORDER n.
3. It is not necessary to describe the phenomenon by an equation in which powers of derivatives occur. ORDINARY DIFFERENTIAL EQUATION OF FIRST ORDER, FIRST DEGREE.
   It is necessary to describe the phenomenon by an equation in which powers of derivatives occur. ORDINARY DIFFERENTIAL EQUATION OF FIRST ORDER AND HIGHER DEGREE.

The foregoing key is by no means an exhaustive list of the categories into which differential equations can fall. However, the above categories include the vast majority of all the equations which have ever been used to describe biological phenomena. Readers desiring a knowledge of all types of differential equations available should consult any text on advanced calculus, or differential equations. (Parke, 1958, has published an annotated bibliography of such texts).

Of the above four broad classes of equations, by far the most commonly used to describe biological phenomena is the ordinary differential equation of first order and first degree. These can all be written in the form

$$M(x,y)\,dx + N(x,y)\,dy = 0,$$

and will be discussed in detail later.

This author is not aware of any instance where equations of first order and higher degree have been used in biology.

Equations of order n have been used occasionally to describe physiological diffusion processes (e.g. Rashevsky, 1948) and analogous phenomena such as population dispersion (Skellam, 1951). However, in such cases partial, rather than ordinary differential equations of order n are used.

By far the most ubiquitously applicable equations for describing biological processes are differential equations of first order, first degree, used either singly, or as systems of ordinary or partial differential equations.

Systems of simple linear differential equations have been used by many writers on theoretical ecology as the basis for their thinking about certain phenomena, particularly competition and predation (Chiang, 1954; Bartlett, 1957; Leslie, 1957, 1958; Lotka, 1925; Kostitzin, 1939; Volterra, 1926). In the one published instance where an effort was made to test such systems critically, Leslie

(1957), using Gause's (1934) data, did not find too good agreement between observed and calculated values. Furthermore, the intra-treatment variances in the data he examined were so great that it would be difficult to show that any remotely reasonable hypothesis did not fit the data. A great many writers have levelled a variety of criticisms against the Lotka-Volterra equations, on both a priori and a posteriori grounds (Andrewartha and Birch, 1954; Beverton and Holt, 1957; Ivlev, 1945; Nicholson and Bailey, 1935; Smith, 1952; Stanley, 1932; Thompson, 1939; Ullyett, 1953; Watt, 1959a, 1960).

There are a number of ways in which we could begin to think about developing a particular sub-sub-model, but this author has found the following procedure helpful.

Suppose that a dependent variable, Y, is under the influence of X_1, and that the mode of influence is subject to modification by X_2 and X_3. We could write

$$Y = f(X_1, X_2, X_3).$$

Now if we wish to begin by writing a general differential equation expressing the effect of X_1, X_2 and X_3 on Y, the first problem is to decide which of the following versions we wish to begin with.

$$\frac{dY}{dX_1} = f(X_1, X_2, X_3)$$

$$\frac{dY}{dX_2} = f(X_1, X_2, X_3)$$

$$\frac{dY}{dX_3} = f(X_1, X_2, X_3).$$

Often it will be much more clear to us at the outset how to particularize one of these than the other two. Also, the final resultant expression will be easier to integrate if the variable with respect to which Y is differentiated initially is the variable which enters into the final integral equation in the simplest way.

To illustrate, if we expect our analysis to lead us ultimately to something like

$$E = \frac{1}{a}\left\{1 - \exp\left[-a\,k\,E\,N\left(\frac{1_{min} + e^{d-fN}}{1 + e^{c-bN}}\right)t\right]\right\}$$

it may be much more clear how to proceed if we begin with

$$\frac{dE}{dt} = f(N, t)$$

than if we begin with

$$\frac{dE}{dN} = f(N, t).$$

When we have decided that we wish to write a general expression for, say, dY/dX_2, we reason as follows. It may be possible to think of the effect of the independent variables (X_i's) as operating on Y via more than one causal pathway. For example (Watt, 1960), population density effects fecundity via probability of mates finding each other, competition for oviposition sites, and other forms of competition. Suppose, first, that an examination of all available evidence suggests four such causal pathways for a specific instance. Let us assign these pathways the general labels A, B, C and D. We understand, then, that

$$\frac{dY}{dX_2} = f(X_1, X_2, X_3)$$
$$= f(A, B, C, D) \text{ and that}$$
$$A = f_A(X_1, X_2, X_3),$$
$$B = f_B(X_1, X_2, X_3),$$
$$C = f_C(X_1, X_2, X_3) \text{ and}$$
$$D = f_D(X_1, X_2, X_3).$$

Some thought will generally make it clear whether A, B, C and D must be treated as addends, multiplicands, or otherwise, such as

$$\frac{dY}{dX_2} = \frac{A + B}{C + D}$$

For example, how to combine A, B, C and D will often be apparent from the following two probability theorems.

The addition theorem: the probability that one out of m events, any two of which are mutually exclusive, occurs is equal to the sum of the probabilities of the occurrence of each event separately.

The multiplication theorem: the probability that two stochastically independent events occur together is equal to the product of the probabilities of the occurrence of each event separately.

In the case of population density regulating oviposition, we could write

$$\frac{dE}{dt} = k\,P\,I\,O,$$

where k is a constant, P the effect of probability of finding mates, I the effect of interference, and O the effect of competition for oviposition sites.

Suppose that we have decided we can write

$$\frac{dY}{dX_2} = A\,B\,C\,D.$$

The next problem is to determine how each of A, B, C and D are governed by X_1, X_2, and X_3. (Note that X_1, X_2 and X_3 need not *all* affect *each* of A, B, C and D).

Suppose the experimenter wishes to determine the effect of X_1, X_2 and X_3 on A. He will sort the data, using cards, to separate out the effects of X_1, X_2 and X_3 as outlined in the previous chapter. After tabulating means, as in Table II, he will plot families of graphs, as in Figs. 3 and 4. He will then make various assumptions about the mode of operation of, say, X_1 on A, and check these by examining the graphs. It is best to check assumptions systematically by asking oneself a series of questions, the answers to which form a logical branched tree, as in Fig. 7. Fig. 7, of course contains only the tree resulting from the following partial list of questions. (Some discussion of logical tree diagrams and other tools in abstract algebra is given by Kemeny, Snell and Thompson, 1957).

(1) Is dY/dX proportional to X?
(2) Is the amount of growth in Y with respect to X compounded? That is, is dY/dX proportional to Y?
(3) Is dY/dX inversely proportional to X?
(4) Does dY/dX approach zero as Y approaches some upper asymptote Y max?
(5) Does dY/dX approach infinity as X approaches some lower value X min?

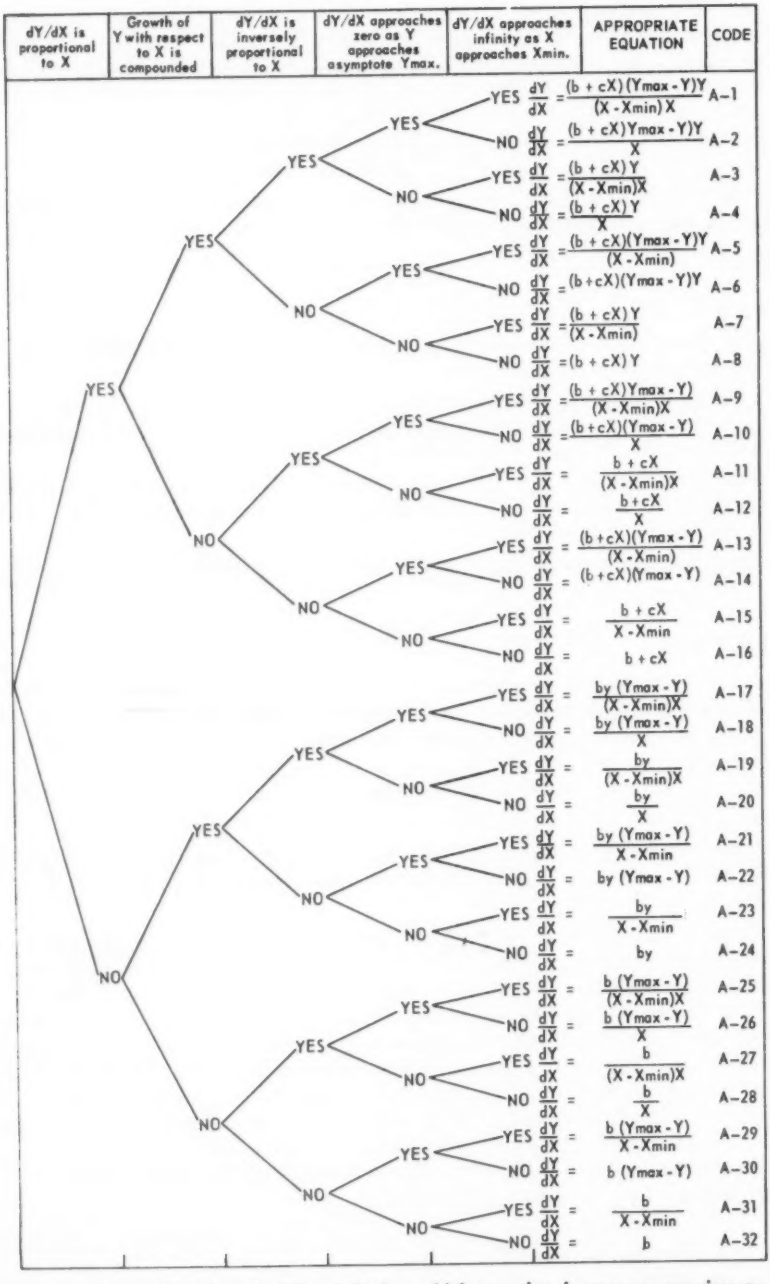
dY/dX is proportional to X
Growth of Y with respect to X is compounded
dY/dX is inversely proportional to X
dY/dX approaches zero as Y approaches asymptote Ymax.
dY/dX approaches infinity as X approaches Xmin.
APPROPRIATE EQUATION
CODE
A-1 dY/dX = (b + cX)(Ymax - Y)Y / (X - Xmin)X
A-2 dY/dX = (b + cX)Ymax - Y)Y / X
A-3 dY/dX = (b + cX)Y / (X - Xmin)X
A-4 dY/dX = (b + cX)Y / X
A-5 dY/dX = (b + cX)(Ymax - Y)Y / (X - Xmin)
A-6 dY/dX = (b + cX)(Ymax - Y)Y
A-7 dY/dX = (b + cX)Y / (X - Xmin)
A-8 dY/dX = (b + cX)Y
A-9 dY/dX = (b + cX)Ymax - Y) / (X - Xmin)X
A-10 dY/dX = (b + cX)(Ymax - Y) / X
A-11 dY/dX = b + cX / (X - Xmin)X
A-12 dY/dX = b + cX / X
A-13 dY/dX = (b + cX)(Ymax - Y) / (X - Xmin)
A-14 dY/dX = (b + cX)(Ymax - Y)
A-15 dY/dX = b + cX / X - Xmin
A-16 dY/dX = b + cX
A-17 dY/dX = by (Ymax - Y) / (X - Xmin)X
A-18 dY/dX = by (Ymax - Y) / X
A-19 dY/dX = by / (X - Xmin)X
A-20 dY/dX = by / X
A-21 dY/dX = by (Ymax - Y) / X - Xmin
A-22 dY/dX = by (Ymax - Y)
A-23 dY/dX = by / X - Xmin
A-24 dY/dX = by
A-25 dY/dX = b (Ymax - Y) / (X - Xmin)X
A-26 dY/dX = b (Ymax - Y) / X
A-27 dY/dX = b / (X - Xmin)X
A-28 dY/dX = b / X
A-29 dY/dX = b (Ymax - Y) / X - Xmin
A-30 dY/dX = b (Ymax - Y)
A-31 dY/dX = b / X - Xmin
A-32 dY/dX = b

Fig. 7. Illustration of method for deducing which equation is most appropriate to describe a phenomenon. A logical "tree diagram" is compounded using a sequence of "yes or no" type questions. The correct model is obtained by following along the appropriate branches. See, for example, Kemeny, Snell and Thompson, 1957.

Many other questions will occur to the reader. A variety of questions with respect to asymptotes and intercepts suggest themselves, for example.

However, the foregoing five questions yield 32 equations which have probably been used more widely than any similar list of 32 which could be assembled.

Equation A-32, the simplest of the 32, integrates to yield $Y = a + bX$, the straight line. Integral forms of some of the other more commonly encountered members of the list are given below.

A-28 $Y = a + b \ln X$

A-24 $\ln Y = a + bX$, or $Y_x = Y_o e^{bx}$

A-16 $Y = a + bX + c'X^2$ (If d^2Y/dX^2 was proportional to X, this would have yielded the cubic $Y = a + bX + c'X^2 + d''X^3$. Other polynomials are analogously derived).

A-20 $Y = aX^b$, or $Y_x = Y_o X^b$

A-22 the logistic,

$$Y = \frac{Y \text{ max}}{1 + e^{c-bx}}$$

A-30 $Y = Y \text{ max } (1 + e^{-bx})$.

This is one way of writing an equation used by Gause (1934) to describe the attack of parasites and predators.

It will very soon become apparent to the biologist working with mathematical models that most of these do in fact have asymptotes. This is to be expected, since most biological processes are limited in some way. For example, a parasite can lay a limited number of eggs, a predator can eat only a certain number of prey per unit time, animals have a limited range of perception and of movement, a biochemical reaction rate is limited by the rate of availability of substrate and so on. Also, the rate of many biological processes does not increase indefinitely in response to increasing variate-values of one or more extrinsic factors. Rather, the process rate increases to a maximum then falls again, under the influence of, for example, increasing temperature. Since reaction rate may change in a complex manner, the influence of, for example, temperature is not describable by parabolas, but more complex functions are necessary. (Ideally, in the case of temperature, a function such as those presented by Johnson, Eyring and Polissar, 1954). The implication of the remarks in this paragraph is that linear multiple regression methods are obviously of no use for building insightful mathematical models intended to mimic the mechanics of complex biological systems.

Once an elementary equation has been chosen to describe the effect of X_1 on A, this equation is integrated (possibly using tables of integrals such as Dwight, 1947) and transformed into a form suitable for testing. There is no need for a statistical test of the relationship, or derivation of parameter values at this stage, since a few simple graphical tests will probably show that the elementary equation is inadequate to describe the phenomenon, if the latter is at all complex. However the best way to proceed will depend on the complexity of the model. The reader should complete this chapter before working on a specific case.

Graphical testing of the validity of hypotheses would be done as follows. Suppose we decided, for example, that A-30 described the effect of X_1 on A while X_2 and X_3 were held constant. Now if

$$\frac{dA}{dX_1} = b\,(\text{Amax} - A),$$

$$-\ln\,(\text{Amax} - A) = bX_1 - \ln A \text{ max},$$

and since we wish, first to see if b is indeed constant, this can be rearranged to yield

$$\ln\left[\frac{A\ max}{A\ max - A}\right] = bX_1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (26)$$

Plotting (26) on semi-logarithmic graph paper (the transformation on the A axis against X_1) will yield a straight line if A-30 correctly describes the relation between A and X_1, X_2 and X_3 being held constant. If a straight line is not obtained, some other elementary equation will be needed in place of A-30. If there is rectilinearity, we proceed to the next step.

This consists of resorting the cards on some other independent variable, say X_2. From (26) we get

$$\frac{\ln\left(\frac{A\ max}{A\ max - A}\right) = b}{X_1}$$

Now if A-30 is adequate as a sub-sub-model describing the influence of X_1 on A,

$$\frac{\ln\left(\frac{A\ max}{A\ max - A}\right) = Z, \text{ say, should yield}}{X_1}$$

a constant when we plot it against X_2. Hence we obtain mean values of X_2 and Z for each of several sub-decks sorted into ranges of values for X_2. If we do not obtain a straight line parallel to the X_2 axis when we plot Z against X_2, we know that A-30 alone is not adequate to describe the effect of X_1 on A, since X_2 modifies the effect. Hence we now need to determine how a term to express the influence of X_2 needs to be entered into A-30. We merely repeat the process done to relate A to X_1. That is, we select an equation on the basis of an examination of the plot

$$\frac{\ln\left(\frac{A\ max}{A\ max - A}\right) = f(X_2)}{X_1}$$

Then an appropriate transformation is used to test the postulated form. For example, suppose we decide that

$$\frac{dZ}{dX_2} = \frac{bZ}{X_2}$$

This can be tested on double-logarithmic graph paper.

It will probably have become clear to the reader by now that we are describing an iterative procedure, for which we could lay out a scheme as follows.

Cycle 1. a. Postulate relation between X_1 and A, using logical branching process as an aid in arriving at hypothesis systematically.

b. Test relation postulated, by making appropriate transformation. (The best collection of techniques on making such transformations which this author has seen is by Running (1917)). If postulate turns out to be inappropriate, repeat a and b until an equation is found which is applicable to relations $A = f(X_1)$.

c. Make another transformation (A′), and graph, to determine if X_2 (or other X_1) has an effect on A′.

Cycle 2. A. Postulate relation between X_2 and A', using logical branching process as an aid in arriving at hypothesis systematically.

The process is repeated, a new cycle to test each hypothesis. It should be noted that there is not necessarily a 1:1 correspondence between the number of cycles used, and the number of X_i's tested, because we will want to test the effect of each X_i on each constant in the equation, as it is built up in stepwise fashion.

When we have finally arrived at expressions for A, B, C and D we are ready for the next step.

This consists of making a final estimate of parameter values, and subjecting the sub-sub-model to appropriate statistical tests.

It should be noted here that a difficulty arises in testing the fit of a highly complex mathematical structure, as the sub-sub-models will be, in general. The equation becomes so flexible due to the large number of parameters it contains, that it is difficult to prove that it does not fit. We have lost a degree of freedom each time we add a parameter. To counteract this, we must try to have as large a number of degrees of freedom as possible. This end is attained by planning the field work so that a large number of places-within-a-year sets of data are obtained. (Planning of field programs will be discussed in detail in chapter 7). Another means of circumventing the aforementioned difficulty is to test component parts of the large equation prior to assembly (insofar as this can be done), then fit all the components together afterwards. This method would only be applicable where we were absolutely sure about the appropriate method of fitting the components together.

If an equation to predict a complex phenomenon can be developed entirely by using data external to a particular series of observations, then tested against these observations, we have a very critical test, losing no degrees of freedom for error at all. Such a procedure was followed by Watt (1955).

Certain books and papers are particularly useful in giving some conception of the difficulty in determining which hypothesis is most appropriate to describe a phenomenon. Hald (1952, pp. 564-570) shows that seven different hypotheses give statistically acceptable fits to a simple set of data, and notes that many others would not have given a significant variance ratio. W. Feller (1940) shows that utterly different growth laws provide reasonably good fits to empirical data, and worse yet, we will get different answers as to which is the best by using different criteria of goodness of fit.

This concludes a general discussion on methods of deriving a deductive model. Prior to considering equations from which we can reason by analogy, we shall turn our attention to various procedures for refining models. Once a preliminary analysis has been completed, and we have a first set of results indicating which variables are important in sub-sub-models and sub-models, and what their structural form should be, we can begin work on refining the model. This step will probably absorb most of the effort expended in data analysis.

The aims of refining the model are twofold: to increase its predictive reliability (decrease the standard error of estimate) and increase the insight into population dynamics yielded by the model. There follow a series of ideas about how these two ends may be attained.

(1) Using the coefficients obtained in regression analysis, obtain values of "Y calculated" for each plot in each year. Then, make up a large graph in which all the calculated Y values are plotted against observed Y values. Such a graph may appear as in Fig. 8.

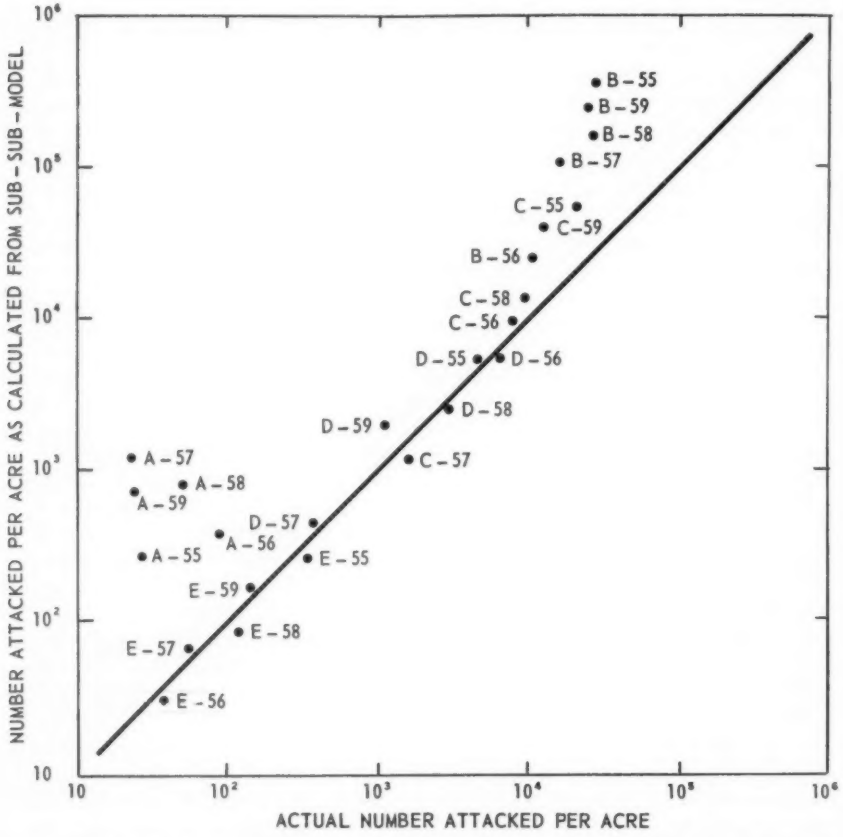

Fig. 8. Values calculated from sub-sub-model plotted against observed values for numbers attacked per acre. Each point is labelled by plot and year in order to facilitate determination of deficiencies in equation.

Each point in such a graph can be labelled as to year and place. A cluster of aberrant points, such as that in the lower left corner may turn out to be due to a weather factor, stand factor, or parasite, say, that needs to be included.

If this device still does not isolate a cause of causes for aberrant points, one can try plotting Yobs. - Ycalc. values against various factors.

Graphs of Y calculated against Y observed, and of Y calculated and Y observed against a particular independent variable, X_i, may reveal places where an incorrect transformation has been used. Various results, with possible interpretations are illustrated in Fig. 9.

(2) It may be necessary to re-examine the way in which certain variables have been defined and coded. For example, suppose we know on the basis of field experience that wind can blow young larvae onto the ground where they will starve at a certain critical stage in their development. However, the regression analysis does not show wind to be important in the form in which we have coded it. Possibly more accuracy is needed with regard to synchronizing wind and migration in the model. That is, the migration period during which larvae are

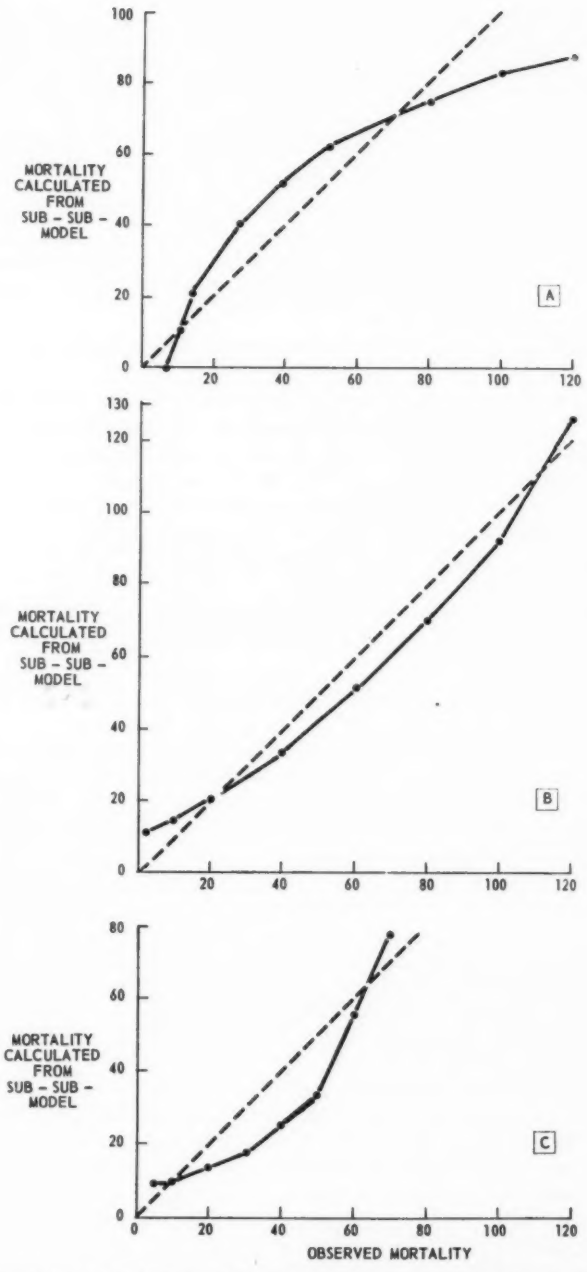

Fig. 9. Simulation of the effect of using inappropriate transformation in sub-sub-model.
A. Data conforming to $\ln Y = bX$ incorrectly fitted to $Y = a + bX$
B. Data conforming to $\ln Y = \ln a + b \ln X$ incorrectly fitted to $Y = a + bX$
C. Data conforming to $\ln Y = \ln a + b \ln X$ incorrectly fitted to $\ln Y = a + bX$

hanging by silk threads is quite short, and we must be careful that the variables for which we have entered values into the regression analysis are those for the maximum wind gusts at precisely this time. Also, the variable we must measure is maximum velocity of wind gusts within an hour, not miles of wind per hour.

(3) Examination of the pattern of regression coefficients of a group of X_i's may show that several are in turn functions of one or more other variables which have not been, but should have been included.

(4) Plotting various X_i's against each other, and families of X_i's on one graph may reveal the need for various interaction terms.

(5) Plotting $Y_o - bX_1 - cX_2$, etc., against X_s may reveal the need for a transformation of X_s. The basis for this transformation will be some differential equation.

(6) Additional variables which were omitted from the first analysis may be brought in.

It will have become clear by now, that what we are in effect describing in this chapter is a large-scale iterative procedure. In essence, we make a first crude guess at the form of the sub-sub-model or sub-model, test this form, and possibly find we have accounted for 40% of the variance in Y. Then we refine the equation somewhat, and repeat the regression analysis, finding this time that we have accounted for perhaps 52% of the variance in Y. Each step in the refining process will reduce the unaccounted for sum of squares somewhat if the refining has been done correctly.

In addition to becoming progressively better by accounting for progressively more of the sum of squares, the model should become progressively better in that it is more analytic.

There are two means by which we can gradually make the model more analytic.

First, by conducting laboratory experiments on physiology, we can obtain information useful in relating regression constants from field data to various physiological parameters. The aim of such work is to isolate a small number of criteria which can be measured to determine ahead of time how a control agent will behave in a natural pest population's ecosystem.

An excellent start has been made on certain subjects which will be of help in explaining efficiency of control agents in terms of fundamental, measurable parameters.

For example, while numerous authors have proposed various equations to describe the effect of temperature on metabolism and growth in insects (Wigglesworth, 1950, pp. 448-454), these will likely soon be replaced by equations which express the effect of temperature in terms of fundamental physical and biochemical constants (Johnson, Eyring and Polissar, 1954). At the same time, there has recently been published a number of books, monographs and papers analyzing the mechanics of insect flight and locomotion: (e.g. Pringle, 1957; Roeder, 1953; Sotavalta, 1947, 1952, 1954). Now the effect of parasite (or predator) population density, P, and host or prey density, N_o, on numbers attacked, N_A, may be given by

$$N_A = PK\left(\frac{1-b}{1-e^{aN_oP}}\right)$$

In this equation, K, a and b are constants. They describe, respectively, the maximum number of attacks each attacker can generate per unit time, the efficiency of search, and the depression of efficiency induced by intra-attacker

competition of various active and passive types. In view of the aforementioned rapidly advancing development of physiological theories, it should soon be possible to replace K, a and b by terms which express K, a and b in terms of fundamental physiological parameters and the environmental factors which regulate their rates of operation.

When equations have been developed in which N_A is expressed as a function of measurable parameters with physiological, and biochemical meaning, certain points should become clear which will aid in search for best possible control agents. For example, it might develop that there is a very simple relationship between the effectiveness of a parasite against a given host, under given conditions of weather and stand type, and the following parameters.

(1) the temperature coefficients in temperature-activity reaction rate equation.
(2) coefficients of searching efficiency, expressing the accuracy of response by an entomophagous insect to sensory cues in the environment (olfactory, chaemotactile, visual, etc.) leading it to its prey or host. (Wright, 1958, for example, indicates how one might go about making such measurements).
(3) weight of insect.
(4) wing-length.
(5) amplitude of wing stroke.

An example of the route by which applied entomology becomes more analytic is provided by work on insecticides. The first step was to test the effects of a great variety of related compounds (e.g. Tahori, 1955). The next step is to relate empirically derived constants to fundamental properties of the control agent, expressed mathematically as parameters which can be measured. In the case of insecticides, one approach to analysis of fundamental properties was by relating insecticidal action to molecular configuration, using Stuart models (Riemschneider, 1954). Such information can provide a simple means of systematizing the search for best control agents.

A second means by which we can make the model more analytic is to incorporate into the equation for Y, independent variables which are assumed to be random variables. That is, we take cognizance of the fact that due to chance, each independent variable will not have exactly the same variate-value at different points in space, even if all measurements are made simultaneously. Instead of predicting a particular value of Y, we predict a probability distribution of Y's.

The best introduction to this subject for the biologist of which this author is aware is the article by Skellam in the book edited by Cragg and Pirie (1955). More mathematical expositions, in order of ascending difficulty and completeness, are by Kemeny, Snell and Thompson (1957), Feller (1950) and Bartlett (1956). Bartlett (1957), Chiang (1954) and Neyman and Scott (1959) illustrate applications to ecological problems.

It may develop in any given subject that a pure mathematical solution to problems formulated as stochastic equations is difficult or impossible to obtain. Recent literature on mathematical model-building is rife with allusions to the difficulty of formulating and solving stochastic models. For example, Neyman and Scott (1959) state, "One of the notable features in the development of real populations is the possible dependence of the birth rate on the degree of local congestion of the population so that an important generalization needed in the theory of clustering is a device, seemingly easy but one we found difficult,

introducing "density dependent" distributions of the size of the litter". Moran (1959) writes as follows, discussing genetical effects of population subdivision. "The situation is entirely changed if we suppose that the population is divided into two parts in which selection operates in different directions. To discuss the theory of a stochastic model of this kind would be a very difficult undertaking. We shall confine ourselves to considering a deterministic model . . . . . ." In such cases, an alternative procedure to developing a stochastic model and solving it mathematically exists. One can build the chance element into the equations in the form of a game to be played on a computer. The game is so designed that it generates the appropriate probability distribution in Y, the dependent variable. Generating a probability distribution of Y by means of a computer is called Monte Carloing, because the process has a physical analogue in the gambling devices at Monte Carlo. Bartlett (1951) used such "gaming" to obtain prey-predator cycles. A more extensive discussion of Monte Carloing is by Meyer (1956).

It is not at all clear at this writing what the ultimate role of stochastic models in biology will be. There are three possibilities.

1. Despite the formidable mathematical difficulties raised by these methods their undoubtedly more realistic mimicking of events in nature, may provide sufficient motivation to gain them wide acceptance and use.

2. The ultimate fate of stochastic models in biology may be foreshadowed by Beverton and Holt (1957). They say, "Considering population change as a stochastic process, in this way, often gives results appreciably different from those obtained with deterministic models (see Bailey, 1950; Bartlett, 1949a; and D. G. Kendall, 1949), especially for the prediction of critical phenomena such as total extinction of the group. It is our belief, however, that, except in particular instances which have been indicated in the text . . . . , the multiplication of effort both in deriving the stochastic equations and in computing them would not have been justified when the standard of accuracy of our data, the complexity of the biotic system with which we are dealing, and the order of magnitude of the expected discrepancies, are all taken into account".

3. Another possibility is that while analytically, stochastic models are difficult to manipulate, their usefulness justifies working with them in some manner, and Monte Carloing will become very important in biological mathematical model building.

This author is inclined to the position that possibility 2 will apply to analysis of phenomena involving large populations, and 3, and sometimes 1 will apply to the analysis of small-population phenomena, in future mathematical model-building in biology.

It would seem to be of interest here, to specify how large errors of estimation are likely to be due to using a deterministic model where a stochastic equation would be more appropriate. Also, the question will often arise, "What is the effect of population size N on the discrepancy between stochastically, and deterministically derived predictions?"

The simplest way to go about answering such a question is to Monte Carlo, using a table of random normal deviates and a desk calcu'ator, or else program an analogous routine for electronic calculator. A small table of random normal deviates is given by Deming (1938, pp. 252-254), and a very large tool for Monte Carloing giving 100,000 normal deviates has been published by the Rand Corporation (1955).

In detail, the procedure is as follows. Readers wishing to follow through development of equations in detail should consult Leslie (1958), whose equations

are referred to here, as examples. Suppose we wish to predict population size N_{t+1} at time t+1, as a function of population size N_t at time t.

(1) Develop an equation relating N_{t+1} to N_t. Leslie (1958) uses

$$E(N_{t+1}) = \frac{\lambda N_t}{1 + \alpha N_t}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (27)$$

where E denotes expectation, or mean, and λ and α are related to constants in the logistic, from which (27) is derived.

(2) Develop or assume a variance corresponding to the mean defined by (27). Leslie uses

$$\text{var}(N_{t+1}) = 2E(N_{t+1}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (28)$$

(Equation (28) follows from the theory of Markov chains; see Kendall (1949) or Bartlett (1956), argument concluding page 70).

(3) Calculate $E(N_{t+1})$, using a given N_t and the appropriate parameters (i.e. an equation corresponding to (27)).

(4) Calculate the variance, σ^2, and from it σ (using an equation corresponding to (28)).

(5) Obtain as many random normal deviates as are desired from the table. These deviates (Δ's), may be positive or negative, and represent fractions or multiples of the standard deviation distant from the mean. That is, the Gaussian deviate .215 corresponds to the number 106 + .215(20) = 110.3, for a normal distribution with mean 106 and standard deviation 20. The deviate −2.354 corresponds to the number 106−2.354(20) = 58.9.

(6) Generate the probability distribution for N_{t+1}, obtaining each value from the relation

$$N_{t+1} = E(N_{t+1}) \pm \sigma\Delta \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (29)$$

By probability distribution, we mean a curve of frequencies of values or ranges of values for N_{t+1}, against N_{t+1}.

Leslie (1958) has already examined the effect of population size on the difference between stochastic and deterministic predictions using Monte Carloing. Predicting a mean value for ten replicates, where $N_0 = 15$, the deterministic model would yield an answer 11% above the correct answer (derived from a stochastic model) for N_{t+10}. Where $N_0 = 300$, however, the deterministic prediction of N_{t+10} was only 2% over the stochastic prediction. This is an oversimplification of Leslie's results, since he considered several types of stochastic models; his paper should be consulted.

In general, where var $[N_{t+1}] = 2E[N_{t+1}]$, the coefficient of variation at t+1 will be

$$\frac{100\sigma}{\bar{x}} = \frac{100\sqrt{2\bar{x}}}{\bar{x}} = \frac{141}{\sqrt{\bar{x}}}.$$

Hence, if one population is R times the other, its coefficient of variation will be

$$\frac{141/\sqrt{R\bar{x}}}{141/\sqrt{\bar{x}}} = \frac{1}{\sqrt{R}} \quad \text{as great as the other.}$$

On the basis of Leslie's (1958) investigations, it does not seem worth while to consider the further analytic complication of introducing random variables into models unless a population has only a few hundred individuals.

This concludes our discussion of methods for refining sub-sub-models and sub-models. We now digress to consider alternatives to least squares methods as means of estimating parameter values in curves. Since most of these methods are used rarely, relative to the method of least squares, we will only mention them, and give references in case the reader might find a need for one of them.

R. A. Fisher's method of maximum likelihood is discussed in most statistical text books. The presentation in Hald (1952) is one which should help make the concept intuitively clear, and Stevens (1951) outlines the procedure for asymptotic curves.

Many types of functions do not lend themselves readily to least squares or maximum likelihood methods, and for these a great variety of approaches have been developed.

Janisch (1927, 1928, 1932) believed that catenary curves might be of great importance in describing biological phenomena which can be understood in terms of two or more components. Catenaries can be thought of as special cases of equations such as

$$y = A_1 e^{a_1 x} + A_2 e^{a_2 x} + A_3 e^{a_3 x} + \ldots . + \ldots . A_n e^{a_n x}.$$

Special methods for estimating parameters in such equations have been developed by Gheury de Bray (1928), Grumman (1936) and Walsh (1919).

Other publications on miscellaneous methods of obtaining parameter values for non-linear functions are by Elderton (1938), Hartley (1948), Hotelling (1957), Nair (1954), Will (1936) and Wilson and Puffer (1933).

Very large, complex mathematical functions can be fitted on electronic computers, using a variety of techniques. The method of differential correction, is outlined by Nielsen (1956, pages 309-313). Box and Coutie (1956) outline an iterative procedure on electronic computers for minimizing the sum of squares of residuals about a fitted hypersurface.

We now turn our attention to the problem of developing the best possible general sub-model that can be structured at present. In order to do this we must first consider which forms of sub-sub-models seem most useful to describe the actions of various factors, on the basis of our present knowledge.

We shall consider appropriate forms for sub-sub-models in the four categories of Andrewartha and Birch (1954): weather, food, pathogens and other animals, and a place in which to live.

The effect of interacting temperature and humidity on growth, metabolism and survival will never be correctly formulated until we have insight into the mode of action of these factors on enzyme kinetics of animals. An encouraging beginning has been made on the development of equations which incorporate such insight (Johnson, Eyring and Polissar, 1954). However, as a stop-gap measure, it may be possible to express the effects of these factors using very simple empirical formulas (Fry and Watt, 1957; Pradhan, 1945, 1946a, b; Watt, 1959b). Any realistic models for the effect of temperature will have to mimic the action of fluctuating temperatures. Also, suitable mathematical description for the effect of temperature and humidity on desiccation and freezing are needed. For the present, the best we can do is to use the following devices as sub-sub-models in the sub-models.

For the effect of fluctuating temperature and humidity on metabolism (growth, fecundity, fertility, activity) let us use the symbol T_M. We will under-

stand by this a function, incorporating as many independent variables and parameters as are needed. Further research will be needed to determine the appropriate form of T_M.

Similarly, we shall use T_F and T_D to indicate apparent mortality due to the effect of fluctuating temperature and humidity on freezing and desiccation, respectively. These symbols describe functions, which to be realistic, must take cognizance of acclimatization, rate of change of temperature, adjustment to cold temperatures, time of exposure, and most lethal temperature to which the animals were exposed (see, e.g. Andrewartha and Birch, 1954; Fry, 1947; Salt, 1950). It is clear from Salt and James (1947) that no function is presently available to express correctly the effect of temperature on freezing in nature, and the same is true for desiccation.

In the same manner, we shall denote the apparent mortality due to wind and rain by W and R.

Presumably food can affect survival in a number of ways. For example, food plant species or variety, age and density are probably all important. Let us denote apparent mortality due to unsuitability of food by F, bearing in mind that F would in fact represent a term as in equation (11).

The effect of pathogens, other animals of the same kind, and other animals of other kinds must be represented by a series of sub-sub-models, one for each species. Each of these sub-models will be basically similar to equation (9). The apparent mortality of pests caused by parasite or predator species i will be denoted by N_{Ai}. The apparent death due to disease by pathogen i will be denoted by N_{Di}, the apparent mortality due to intraspecific competition by N_C, and due to interspecific competition by species i as N_{Ci}. Intraspecific competition will include not only such effects as starvation due to crowding, and cannibalism, but also dispersal on to unsuitable terrain caused by excessive local density (for mathematical formulation see Skellam, 1951, and Dempster, 1957).

We shall use the symbol H to denote apparent mortality caused by unfavorable characteristics of the habitat. This category will include such factors as tree dispersion within stand, and isolation of stand. It governs the probability that the organism will be killed due to shortage of favorable niches in time of need or emergencies. "Favorable niche" here means a niche suitable to fall on, to hide in from enemies, a niche which smells unattractive to enemies, or a niche which minimizes chances of desiccation and freezing, etc.

Let us now determine how a generalized sub-model incorporating the aforementioned factors should be structured, bearing in mind the principles of structuring outlined in chapter 3.

First, let us assume that all N_{Ai}, N_{Di}, N_C and N_{Ci} expressions are governed by the effect of fluctuating temperature and humidity on the metabolism of either or both of the effector and effected species. Hence we shall write each of these as, for example, $N_{Ai}\,(T_M)$.

Let us assume that two of three attacking species operate mutually exclusively. In line with equation (5), their effect can be expressed as

$$\frac{N_{A1}\,(T_M) + N_{A2}\,(T_M)}{N_{III}},$$

where N_{III} refers to the number of pest larvae present at the beginning of the third instar.

Let us also imagine that the third attacking species, and a disease, operate independently with respect to each other, and the first two attacking species. Following equation (4), we now have a general sub-model of form

$$S_{III} = \left\{1 - \left(\frac{N_{A1}(T_M) + N_{A2}(T_M)}{N_{III}}\right)\right\}\left\{1 - \frac{N_{A3}(T_M)}{N_{III}}\right\}\left\{1 - \frac{N_{D1}(T_M)}{N_{III}}\right\} \quad \ldots\ldots\ldots\ldots (30)$$

Now if there is some mortality due to desiccation and wind, to exemplify the operation of weather factors, death due to each of these would in turn be modified by the accessibility of protective niches. Hence we must denote these by T_D (H) and W (H). It can not be emphasized too much that H would itself represent a complex mathematical expression, as would each of T_D and W. Symbolism such as T_D (H) therefore means that a complex function (e.g. T_D) is itself a function of another complex function, H. Not enough research has been done to particularize with certainty the form of any of these functions.

Combining the factors in (30) with the above two sub-models, we arrive at (31) as the final most appropriate form for a sub-model incorporating these sub-sub-models.

$$S_{III} = \left\{1 - \frac{T_D(H)}{N_{III}}\right\}\left\{1 - \frac{W(H)}{N_{III}}\right\}$$
$$\left\{1 - \left(\frac{N_{A1}(T_M) + N_{A2}(T_M)}{N_{III}}\right)\right\}$$
$$\left\{1 - \frac{N_{A3}(T_M)}{N_{III}}\right\}\left\{1 - \frac{N_{D1}(T_M)}{N_{III}}\right\} \quad \ldots\ldots\ldots\ldots (31)$$

The typical product of a few seasons' field work will be a mixed model, containing some terms derived through regression analysis, and some derived from combined deduction and induction. Such will arise when we can develop structural terms for some factors, but not for others. The motive for handling an equation in this form is to derive a means of using information gained about some factors as an aid in pin-pointing which other types of factors and factor interactions need to be incorporated into sub-models.

For illustrative material, let us assume an equation for egg survival, S_E, where there would be in the final deductive model five sub-sub-models, all interacting at random with regard to each other, as in equation (4). Equation (32) is hence merely a generalization of (4).

$$S_E = \left(1 - \frac{M_1}{N_E}\right)\left(1 - \frac{M_2}{N_E}\right)\left(1 - \frac{M_3}{N_E}\right)\left(1 - \frac{M_4}{N_E}\right)\left(1 - \frac{M_5}{N_E}\right) \quad \ldots\ldots\ldots\ldots (32)$$

Now suppose M_1 and M_2 can be evaluated, using, for example, biological techniques outlined by Miller (1955) and equations outlined by Watt (1949a). Suppose, however, M_3, M_4 and M_5 can not be evaluated in terms of structural sub-sub-models. The problem is to determine, say, which of these three is most worthy of expenditure for a medium-sized research program, and hence which is of most potential usefulness as an auxiliary control agent. Also, we want to determine if it is worth while to undertake field studies on any possible interactions of the three remaining factors.

Having postulated the form (32), we can rewrite this as

$$\frac{S_E}{\left(1-\frac{M_1}{N_E}\right)\left(1-\frac{M_2}{N_E}\right)} = S_E' =$$

$$1-\frac{M_3}{N_E}-\frac{M_4}{N_E}-\frac{M_5}{N_E}+\frac{M_3M_4}{N_E^2}+\frac{M_3M_5}{N_E^2}$$

$$+\frac{M_4M_5}{N_E^2}-\frac{M_3M_4M_5}{N_E^3} \qquad (33)$$

Equation (32) can be transformed into (33), which will be recognized as a particular case of (10), and therefore lends itself to the multiple regression approach discussed in chapter 4. We may use the analysis of variance in conjunction with regression analysis, to determine which of the effects M_3/N_E, etc., and interaction effects such as M_3M_4/N_E^2 are worthy of further study.

6. METHODS OF USING MODELS

Various motives for developing a mathematical model of pest population dynamics were given in chapter 2. Let us now reconsider each of these aims, and determine how the completed model could be used to attain them.

It was stated that a model which is sufficiently analytic (i.e. which incorporates enough insight into the mechanics of operation of each factor) can be used to find optimum control agents. Suppose, for example, that the basic equation for effectiveness of a parasite is

$$N_A = PK\left(1-e^{-aN_oP^{1-b}}\right)$$

Suppose, further, that K is a function of body size (e.g. Miller, 1957), and temperature and humidity (e.g. Atwal, 1955; Birch, 1953; Burnett, 1954, 1956; El-Sawaf, 1956; Menusan, 1934; Payne, 1933). For purposes of illustration, let us suppose that the best simple function that can be devised in our present state of knowledge, for K, is

$$K = (r+sL)\left(e^{-dx_1^2}+He^{-fx_2^2}\right)$$

where L is the insect's body length, r, s, d and f are constants characteristic of the species, H is the relative humidity, and x_1 and x_2 are the differences between any observed temperature and two different optimum temperatures.

Similarly, a, which measures searching efficiency of the species can be expressed in terms of sensory-cue following accuracy (A) and, say, four physiological and biochemical parameters of aerodynamic work output (F_1, F_2, F_3, F_4).

Finally, b, which measures depression of parasite and predator searching efficiency caused by various forms of intraspecies competition pressure, can be expressed as a function of anti-superparasitism discriminatory ability, D, A, and say, F_1, F_2, F_3, and F_4. Inclusion of F_1, F_2, F_3 and F_4 follows from our assumption that increased flight power can overcome the deleterious (superparasitism or superpredation) effects of increased attacker density.

Bearing these points in mind, (9) may be rewritten as the general equation.

$$N_A = P\left[(r + sL)\left(e^{-dx_1^2} + He^{-fx_2^2}\right)\right]\left[\frac{}{1-e^{-a\,(A,\,F_1,\,F_2,\,F_3,\,F_4)\,N_oP^{1-b\,(D,\,A,\,F_1,\,F_2,\,F_3,\,F_4)}}}\right] \quad \ldots\ldots\ldots\ldots (34)$$

Now once a structural equation analogous to (5) has been determined for each of parasites, predators and diseases, it will be clear precisely what it is about each of these three types of agents which governs their effectiveness.

For example, it would be apparent which metabolic and aerodynamic features make for a most effective searcher, and optimum control could be obtained with the species that combined these characteristics to the highest degree. Obviously the matter is somewhat more complex than this discussion would indicate. To illustrate, at endemic pest levels, the best parasites and predators are thorough searchers, whereas under pandemic conditions, very high biotic potential is the salient desideratum.

Equations such as (34), a typical sub-sub-model, satisfy the second need for mathematical models, in that they clarify research objectives. For example, in order to formulate, and obtain parameter values for (34), one would need to obtain readings from a large number of parasites in each of several species on

(a) the length-humidity-temperature-fecundity cause-effect system
(b) the relation between searching efficiency and the factors which govern it
(c) depression of searching efficiency and its dependence on parasite density, and the factors peculiar to each species which modify the effect of parasite density.

It follows as a corollary statement from the preceding paragraph that another need for models, clarification of research objectives is also satisfied by sub-models such as (34). In order to know which parameters of a parasite or predator, say, need to be measured, we must have done fundamental research on an array of species to determine which are relevant governors of the phenomena under consideration.

Such research would combine two types of work. First it would be necessary to measure constants such as a in the field for a great many species. Second, it would be necessary to conduct physiological, biochemical and behavioristic studies in the laboratory to analyze mechanisms and discover what factors determined a.

Once such research has been done, instead of evaluating parasites and predators from all over the world merely by releasing them in the area where they are needed and awaiting results of this procedure with mixed hope and trepidation, one would need to collect and release only those parasites and predators which conform to requirements determined through research such as that aforementioned.

The other need for models, and the means by which it can be satisfied, has been deliberately deferred to the end of this discussion. Obtaining values of the various manipulatable variables that result in minimization of the trend index constitutes a difficult mathematical problem. Furthermore, the methods which are available for solving such extremum problems have as yet seen essentially no use in biology. (Some of them are within the domain of a new branch of mathematics developed largely in the last seven years, "mathematical programming", a quite different subject than computer programming).

Classical and modern mathematics has provided us with seven possible approaches to the problem of determining how to maximize or minimize some value. Which of these we select in any instance depends on either, or both of, the fundamental nature of the problem and the way in which we choose to formulate it. We shall now discuss the seven methods very briefly in turn, pointing out in each case what characteristics our formulation of the problem must have in order for the method to be applicable.

Some of the methods seem to this author to be patently inapplicable to the minimization of a pest's trend index. However, they are included to add to the intelligibility of the discussion, since all seven methods are to some degree conceptually or methodologically related.

1. Suppose we wish to minimize the trend index, $T_{t;t+1}$, and we are completely free to manipulate any variables which enter into the model for T. Let P_i denote the ith parasite or predator, D_i denote the ith disease organism, W_i the ith weather factor, and A_i the ith stand factor. Let us suppose further that trend index may be expressed as some function.

$$T_{t;t+1} = f(P_1, P_2, \ldots P_n, D_1, D_2, \ldots D_n, W_1, W_2, \ldots W_n, A_1, A_2, \ldots A_n) \quad (35)$$

Now we can find the values for each of the independent variables which minimize $T_{t;t+1}$ as follows, provided (35) has certain properties. First, it must be differentiable with respect to each of the independent variables, and second, it must have a minimum, or various local minima which can be found by iterative exploratory procedures.

We obtain the system of partial differential equations

$$\begin{aligned} \frac{\partial T}{\partial P_1} &= f'_{P_1}(P_1, P_2, \ldots A_n) \\ &\cdot \quad \cdot \\ &\cdot \quad \cdot \\ &\cdot \quad \cdot \\ \frac{\partial T}{\partial A_n} &= f'_{A_n}(P_1, P_2, \ldots A_n) \end{aligned} \quad (36)$$

Then each equation in (36) is set equal to zero, and all are solved simultaneously for the appropriate values of P_1, P_2 etc.

Two objections to this model are apparent immediately. First, man does not have control over all the relevant independent variables, and second, the simultaneous solution of all the equations in set (36) may raise formidable mathematical difficulty.

However, a principle of operation underlying all subsequent methods is exposed. Methods analogous to 1 can tell us, in principle, what values each variable should take to minimize a pest's trend index. This implies that the value zero for any variable can arise, and hence such methods can, in principle, tell us if a control agent should not be used at all. This would be true if the deleterious effects of an agent on other agents more than nullified its positive effect in decreasing survival, fecundity or the proportion of females in the pest species (Watt, 1960). It could also happen, for example if an insecticide reduced survival but increased fecundity, say (e.g. Hueck et al, 1952; Knutson, 1955; Kuenen, 1958; Watt, 1960).

2. Let us now make the problem statement somewhat more complex (and at the same time more realistic). Suppose that we wish to minimize the trend index T, which is a function of the variables mentioned in 1, plus N_t, the number of pests

alive just prior to oviposition in year t. Let us assume that we have no control over N_t, or the W_t (weather factors). That is, in any year t we expect to find that of the several independent variables which regulate $T_{t,t+1}$, some of these are beyond our control. This is CASE C of Watt (1955, page 280). In such instances, the set of equations corresponding to (36) will contain one equation for each independent variable over which man has control. The other independent variables will be treated as constants throughout the set, the particular values used being those encountered in year t.

We might speak of the problem posed this way as being one of finding the minimum of a function subject to point constraints.

3. The problem formulation in 2 still isn't very realistic, because in addition to point constraints on solution, there are generally other constraints which are somewhat more elaborate. For example, suppose we consider chemical control. If dosages sprayed become too great, and if there are too many spray treatments per season, a number of consequences follow. First, the chemicals become prohibitively expensive, and overhead wipes out any profit in growing the apples, wheat, trees or other crop. Second, there are deleterious side effects on other fauna. Third, legal tolerances for residues of insecticides may be exceeded. The same types of considerations apply to cultural control or biological control. In the case of forest management, it appears that the trend index of spruce budworm, for example, could be minimized if forest stands were sufficiently thinned (Morris et al, 1958). However, if the stands were thinned considerably, the resultant forest might constitute an inefficient use of available soil, and also the thinning process itself could be excessively expensive. In the case of introduced predators, parasites, and diseases, too many are prohibitively expensive, and too few are ineffective and unduly prone to extinction. The chance of a population becoming extinct is given by

$$\left(\frac{d}{b}\right)^N,$$

where b and d are birth and death rates and N is initial population size (Bartlett, 1956).

If we now consider the various types of restrictions exemplified above, we see that they are interrelated. The maximum amount of money we could spend on releasing imported parasites and predators is dependent on the amount we spend on other control measures. That is, the total contribution all control agents make to overhead must not exceed a certain value. Similarly, some of the deleterious side effects of different parts of the spray schedule may be additive. Instead of postulating a series of point constraints as in 2, we can formulate the extremum problem more realistically by assuming equational constraints.

Here, the basic equation is

$$T_{t,t+1} = f(N_t, P_1, P_2, \ldots A_n) \quad (37)$$

but it is subject to constraints of form

$$p_1P_1 + p_2P_2 + \ldots\ldots c_3C_3 = 2700 \quad (38)$$

Here p_i denotes the cost per parasite or predator of releasing imported entomophagous insects, C_3 represents the cost of applying the third chemical treatment to a unit, say an acre. The maximum allowable cost of protecting an acre, say, is set at 2700. dollars.

If we pose the problem as in (37), and express the constraints as in (38), the

problem can likely be solved using Lagrange multipliers, which are discussed in any text on advanced calculus.

The three formulations of the problem of minimizing a pest's trend index already given can be handled by classical methods of calculus. The next three methods all require some modifications or extensions of classical methods. They all have another feature in common. If the reader reflects a moment, he will see that the constraint in (38) does not express the idea in the paragraph preceding (38). There it was stated, ". . . the total contribution all control agents make to overhead *must not exceed* a certain value". Equation (38) states that the total contribution must *equal* a certain value. This difference, which might appear trifling, and the associated inadequacy of constraints expressed as in (38) constituted the motivation for developing modern methods of handling extremum problems. Now if the reader considers the matter, he will see that almost all constraints on control practices must be expressed as either

$$P_1 \leqq x, \text{ or} \quad (39)$$

$$p_1P_1 + p_2P_2 + \ldots\ldots c_3C_3 \leqq 2700, \quad (40)$$

rather than

$$P_1 = x, \text{ or}$$

$$p_1P_1 + p_2P_2 + \ldots\ldots c_3C_3 = 2700.$$

It is never necessary that exactly x of a control, or of all controls, or x dollars worth of all controls must be used. Rather, it is true that some amount equal to or less than x must be spent, or that an amount of DDT equal to or less than x must be sprayed, or a number of predators per acre equal to or less than x, but equal to or greater than y must be released.

We will now consider the three methods of solving extremum problems which formulate the problem using inequational constraints, as in (39) and (40). All these methods are referred to as "mathematical programming". The first to be conceived was "linear programming", developed by George Dantzig in 1947, and published by Koopmans (1951). The methods have largely been applied to problems in operations research and econometrics occuring in business, industry, and analysis of national economy.

4. If we are willing to assume a linear basic equation (as in (37)), and linear inequational constraints, the minimization of trend index can be regarded as a problem in linear programming (Dorfman, et al, 1958; Gass, 1958; Kemeny et al, 1957; Koopmans, 1951; Vajda, 1956). However, this is again not a realistic way in which to formulate our problem. The one fact of which we can be most certain is that the trend index equation will not be even approximately linear. Many of the sub-sub-model processes will decrease in rate as asymptotes are reached, as in equation (9).

5. Where basic equations are non-linear, in some cases it will be possible to use a method analogous to linear programming called non-linear programming (Arrow et al, 1958; Dorfman et al, 1958; Kuhn and Tucker, 1951). Not enough is known yet about the possible form of sub-sub-models, sub-models and models to know if this general approach will be applicable. However, its assumptions are markedly more realistic for this problem than those of the previous four formulations.

6. In all the cases to this point, we have assumed that the problem could be formulated in the most useful and realistic fashion by considering only one time

interval, t to t + 1. However, there is a quite different way of looking at insect pest control. This will become apparent if we reconsider equation (37)

$$T_{t;t+1} = f(N_t, P_{1,t}, P_{2,t} \ldots\ldots\ldots A_{n,t})$$

Now as a result of setting certain of the independent variables equal to specified variate-values, and as a result of the remaining independent variables (weather, etc.) taking certain values, in year t, (37) yields a value for $T_{t;t+1}$. But this value in turn defines N_{t+1}, since from equation (1),

$$N_{t+1} = T_{t;t+1} N_t \ldots\ldots\ldots\

matics. Furthermore, the specific problems they have been shown to be capable of handling are extremely simple relative to problems that may arise in the real biological world. Suppose that an analytic solution of the extremum problem for pest's trend indices is impossible, due to insurmountable mathematical difficulties encountered. What do we do then?

7. Fortunately, there is another avenue of attack available to us. When Bellman (1957) talks about the formidable computing job of calculating minima or maxima for $T_{t;t+n}$, he is referring to the case in which one would simply search every cell of the mn-dimensional space within which $T_{t;t+n}$ occurs to locate inflexion points. Now as Box and others have pointed out recently, it is not necessary to locate maxima by such a prodigal computing effort. (Box, 1954; Box and Wilson, 1951; Box and Youle, 1955). Computers can be used to locate maxima in n-dimensional surfaces by sequential maxima-hunting strategies (Box and Coutie, 1956). Box's ideas have been applied in a biological study by Hader et al (1957) and Moore et al (1957).

7. THE FIELD PROGRAM DICTATED BY THE MODEL

Suppose a specific insect species is a pest, and administrators feel that a field program is required to obtain data useful in controlling the species. How should a program be designed to collect data satisfying input desiderata of mathematical models as outlined herein?

Above all, the researcher should be realistic in planning a field program, and tailor the amount of work done to the aims of the study, the complexity of the phenomenon, and the accuracy requirements. Too much or too little information is wasteful.

"If we really don't need much information, or a high degree of predictive reliability, it is wasteful to use a research plan based on the more complex types of models, since these require an expensive, comprehensive program of data collection. Conversely, if we do require a high degree of predictive reliability, and a lot of information about the population dynamics of the . . . species, we can not expect to achieve these ends unless we are prepared to maintain a costly and extensive field program. Furthermore, it should be emphasized that the complex, expensive program will probably have to be kept in operation for at least 20 years, before it justifies its existence, in order that enough data can be collected to sort out the effects of all the independent variables, and study the pattern of the way in which they interact. Finally, the whole program will have to collect data at a sufficiently high rate that the results of the work will not defy analysis due to small-sample error." (Watt, 1959).

While the material in this monograph can not be applied completely until 10 or more years of data are in, information from the beginning should be collected so that it is possible to apply such methods of analysis. Otherwise, the study may have represented many years of wasted public moneys; any statistician who has consulted for field biological groups at the end, rather than the beginning of such studies can name many instances of waste.

Let us consider first the effect of complexity of the phenomenon on the necessary volume of replication. It is clear that there is indeed a real difference in the number of factors regulating pest populations. This number may be very large as in the case of spruce budworm (Morris et al, 1958) and grasshoppers (Spencer, 1958), or very small, as in the case of the oystershell scale in the Annapolis Valley. Given a rough estimate of the number of factors, we may arrive at an approximate estimate of the number of units of data required as

follows. (Here, a data unit is meant as a series of measurements on a plot for one year, each measurement to be based on a set of replicates).

Suppose we were trying to define the mathematical form of a relationship between a dependent variable Y and one independent variable X_1. Any curve, from a circle to a straight line, can be drawn through two points, and clearly we need eight or more points to discriminate between alternative hypotheses about the form of $Y = f(X_1)$. Now if we had two independent variables, clearly the number of sets of values for Y, X_1 and X_2 we would need is closer to 8^2 than 8. Hence, for a very elaborate model involving n independent variables, the goal we should be aiming for is about 8^n sets of data.

However, not only is the number of sets of measurements critical, but also the distribution in space and time of sampling effort. Sampling plots should be so chosen as to cover the widest possible variety of combinations of values for the complete set of independent variables. That is, while we are committed, in a large scale exploration of interacting natural phenomena, to work within the conceptual statistical framework of a sample survey, rather than a designed experiment, we try to choose plots so we have, insofar as possible, a factorial experiment.

For each plot, sampling should be continued for as many years as possible.

Also, there is no point in studying three plots for three years, then dropping one and starting another, then similarly switching the other two the following year. The study must have continuity, in that the same plots are studied each year. Otherwise, whoever analyzes the data finds himself in the logically untenable position of trying to analyze an effect at one site as a function of cause at another site the previous year. Some scientists have an unfortunate habit of allowing their curiosity to run away with their common sense. By this I mean that as field programs run for more and more years, the staff becomes curious about what they would find if they sampled other locales. Finally, they become so burdened with work that they have to discontinue sampling at the locales where they initiated the program. The result is that there is, at termination of 20-years work, no more than a few years sampling data at any one site. This author has seen and heard of such tragically wasteful calamities befalling a great number and variety of studies. As a general rule, if insight into mechanisms is the aim, intensive work at a few sampling sites is much more rewarding than extensive work at a great many.

Very careful field work will be necessary to ensure that all relevant factors are selected for study. A serious error has been committed if either not enough, or too many factors are studied over several decades. The former will prevent the standard error of estimate from being small, the latter will be wasteful.

A major aim of field work should be to maximize the useful information output for a given amount of field work expended. This can be done by using the analysis of variance to determine where the bulk of the sampling error originates then designing the field work accordingly (Bancroft and Brindley, 1958; LeRoux and Reimer, 1959; Morris, 1955). Bennett and Franklin's (1954) discussion of models for obtaining variance components is particularly detailed. Once components of variance attributable to each level of an experimental or survey hierarchy of levels have been estimated, these can be used in cost functions to either minimize the total cost of obtaining a given variance or minimize the total sampling variance for a given cost (Bancroft and Brindley, 1958; Bennett and Franklin, 1954; Hansen, Hurwitz and Madow, 1953; LeRoux and Reimer, 1959; Morris, 1955).

It has been deemed unwise to present the theory and examples of its application to actual entomological situations here. Any explanation less than a very complete one would be misleading, since the estimation of variance components is an excellent example of a field in which a little knowledge is dangerous. However, a complete explanation would add a very long section to this work, and would merely repeat what has been thoroughly discussed already by Bennett and Franklin. This very excellent book is wholeheartedly recommended as a most suitable, comprehensive introduction to the theory of analysis of variance for the biologist.

Three types of pitfalls await the entomologist intending to use elaborate applications of the analysis of variance, and he should note these rubrics carefully.

A. The analysis of variance will not yield accurate estimates of variance components if certain of the assumptions on which the method is based are badly violated. Transformations may be necessary to correct this situation. The four basic assumptions underlying the method are as follows.

1. The effects of, say, plots, trees within plots, and levels within trees must be additive.
2. Variances and covariances must be homogeneous (the same from class to class, etc.).
3. The distribution must be normal.
4. The residuals must be random.

Violation of assumption 3 is not generally important. Whether or not it is most important to correct for violation of 1 or 2 is a contentious point, but where analysis of variance is being done primarily to obtain estimates for components of variance, it is most important to comply with assumption 1. In general, the transformation log $(x + 1)$ will be suitable for entomological data. This follows from the fact that population growth is geometric, not arithmetic and different counts in different places result from different geometric growth rates. An excellent discussion of transformations suitable for the biologist, is by Quenouille (1950, chapter 8). Tukey (1949) outlines a test for non-additivity and Hamaker (1955) illustrates its use. One of the tests mentioned by the above authors should be applied before the decision is made to make any particular transformation. Also, some judgement and experience will be helpful in determining if any transformation at all is worth the effort. In general, transformations are not worthwhile if (a) the data do not cover a very wide range of values, and (b) the sampling scheme developed from variance components using cost functions employs a crude assignment of effort to different cases within one hierarchical level. For example, if we derive a sampling scheme which states that we sample 2 branches per tree and 8 trees per plot, these numbers scarcely justify a very high number of significant digits in the variances from which they are calculated. Some experimenting with dummy sets of figures will show how much accuracy in estimates of variance components is worthwhile, and hence, whether transformation is necessary.

B. The biologist should note carefully that different computational formulas for the analysis of variance are required, depending on the nature of the sampling scheme with respect to the population under study. The first mode of classifying models for the analysis of variance divides them into three categories, convention-

ally referred to as models of type I, II and III. Their definitions follow, in terms of the simple equation,

$$x_{i\alpha} = \mu + \beta_i + \epsilon_{i\alpha} \quad ,$$

in which $x_{i\alpha}$ represents any observation α in class i, μ is the mean for all classes. β_i is the additional effect associated with class i, and $\epsilon_{i\alpha}$ is a random variable representing the variation of the $i\alpha$th observation from $\mu + \beta_i$, the mean value for the ith class.

I. If β_i is regarded only as the mean value for the ith class examined, and no information is to be inferred from all the β_i to means of unmeasured classes, model type I is appropriate.

II. If the β_i are regarded as a random sample from an infinite population of possible values of β_i, we can estimate the variance of the population of all possible β_i's.

III. If the β_i are regarded as a random sample from a finite population of β_i's, model type III is adopted.

C. A second mode of classifying models depends on whether we are dealing with crossed or nested classifications, or variants or mixtures of these. If we have four levels in a tree, and within each of these sample four compass quadrants, so that the north quadrants are on top of each other, we have a $p \times q$ crossed classification. It will probably be appropriate in the analysis of variance model to introduce an interaction term for such cases. That is, we expect the effect of a branch in the north quadrant to be different depending on which level the quadrant is in. Now if within the ith level, of p levels and jth quadrant, of q quadrants, we sample n branches, this is not a crossed classification of form $p \times q \times n$. This is because the αth branch in the ijth quadrant is no more similar to the αth branch in the $(i + 1)$ $(j + 1)$ th quadrant than it is to the $(\alpha + 1)$th branch. We have only taken a random sample of branches from each quadrant, and here we have a nested classification, within a crossed classification.

To summarize, before using the analysis of variance to obtain variance components, the biologist should ask himself these questions:

(1) Do I need a transformation, and if so, which one? If there does seem to be a need for one, does the effort involved in converting my data seem justified by the increased accuracy of the results?

(2) Which model should I use, I, II or III?

(3) Am I dealing with nested, crossed or mixed classifications?

The reader should be reassured at this stage that all problems will not require intensive effort by large teams for indefinite periods of time. After a few years work has been done, it may become clear, that one, or a small number of factors are "key" factors which regulate the pest's trend index from year to year (Morris, 1959).

A relatively inconspicuous factor may be regulating the population, if it is the only one which is density-dependent. The fact that an important factor may not account for much of the average year-to-year mortality emphasizes the need for a relatively thorough study of the relevant factors in the exploratory phase of a field program. If only the factors contributing to the bulk of the average year-to-year mortality are measured, the actual regulatory, or "pacemaker" independent variable may not be studied at all.

Some hypothetical figures will illustrate how the key factor may at first examination seem quantitatively insignificant. Suppose that non-key factors

cause an average of 94% mortality from year to year while a key factor causes 0 to 6%, depending on the pest density. Then, where 98% mortality allows a population to maintain itself at equilibrium, 94% + 2% yields a trend index of 2 (from (100-96)/2), 94% + 4% yields a trend index of 1, and 94% + 5% yields a trend index of .5.

A very recent review by Morris (1960) stresses that the sampling plan should be determined by the precise aim of the sampling program.

A common problem in insect sampling seems worthy of special consideration here. The effect of n different insecticides on the pests on n different varieties of, say, potatoes is investigated by laying out a Latin square of n^2 plots. The conceptual unit of study is the plot. However, we can obtain a great deal more information about the experimental results by analyzing data on individual plants separately, in addition to performing the usual analysis of variance. The following argument will explain why.

Suppose we have 100 plants per plot, and a complete census of a plot shows that it contains 300 phytophagous insects of species A, yielding an average of 3 per plant. This mean of 3 could represent a great variety of biological situations, but to make the point, we shall consider only three highly improbable possibilities.

(1) 3 of A on each of the 100 plants.

(2) 30 of A on each of 10 plants.

(3) 300 of A on one of the 100 plants.

Ultimate total plant destruction will be greatest in case (1), because here there will be virtually no intraspecific competition and the food consumption per individual A will be at a maximum. However, 3A per plant may not be enough phytophagous insects to erode the homeostatic capability of the plant to adjust for damage (assuming such capability exists). In case (2), the effectiveness per A will be lower due to intraspecific competition, but even so, 30 individuals per each of 10 plants may be enough to destroy 10 plants. In case (3), the effectiveness per A will be at a minimum due to intraspecific competition and the ultimate total destruction will probably be considerably less than in case (2). The foregoing argument is most valid for completely sessile organisms, least valid for highly mobile organisms.

The point is that the plant is a more meaningful unit of study in such cases than the plot. The best method of analysis would appear to be to set up a regression equation in which plant destruction for any plant is expressed as a function of the organisms present on the plant prior to treatment, and treatment.

This argument can be amplified by considering the interaction of chemical treatment and entomophagous insects. A mean of, say, one predator per plant in a plot has very different biological significance depending on how the predators are distributed amongst plants. The predators, as the phytophagous insects, are subject to intraspecific competition pressure. All other things being equal, 100 predators will accomplish much more predation if they are sparsely distributed than if they are all aggregated in a small number of clumps.

8. POSSIBLE CRITICISMS OF THE MODEL APPROACH

A number of criticisms can be levelled at the model approach to insect pest control, in principle. Three of these are of particular importance.

The first criticism is that the aims of this approach are much too ambitious, and indeed, are simply unattainable. The answer to this point is that it is much too early to know if the ends are attainable or not, and furthermore, it really does

not affect the answer as to whether the effort is worthwhile. The inherent value in this approach is that in addition to a single large payoff at the end of the field research program (optimum control), there is a payoff for each step in research completed (improvement in control). Also, not only is the yield from each part of the program independent of the yield from the whole program, but it is also partially independent of the yield from other parts of the program. That is, additional insight into the factors which govern parasite effectiveness is valuable, even if we do not concurrently gain additional insight into the factors governing insecticide effectiveness, and vice versa. It is also noteworthy that it is just as expensive to have N men working on various bits of a model, and hence in such a manner that the bits can not ultimately be integrated, as it is to have them working cooperatively on an integrated program.

The second major criticism that can be raised against the use of models in insect pest control is that of Thompson (1939, page 380). His entire discussion is well worth very careful study, but one paragraph gives the general flavour of his argument.

He states, "It is difficult and indeed unwise to set any limits to the usefulness and practical application of mathematics and especially of an instrument as powerful as the infinitesimal calculus; but so far as the writer can see, there is, in the last analysis, no way of developing a method that can reduce to manageable form the appalling complexity of natural factors or smooth out their unpredictable irregularities. When the mathematicians can produce a system enabling us to predict the climatic conditions next year, next week, or even tomorrow, and reduce their fluctuations to mathematical laws, we may begin to talk about mathematical theories of natural populations. It does not seem that they are likely to do this very soon."

Now this point is very cogent, important, and relevant, and should be carefully considered. An instance of how important, quantitatively, unpredictable weather or other changes can be is given by a life table on spruce budworms (Morris and Miller, 1954, p. 288). Here the expected trend index was 36%, but the actual index was 330%, due to invasion by adult moths.

It seems to this writer that whether or not Thompson's point implies that model construction is a waste of time depends on what we want to do with models. If they are for precise, short-term, local prediction, then they are doomed to be grossly disappointing because of chance elements. If they are for long-term average predictions over wide areas, they will appear to be considerably more useful. (See, for example, Miller, 1959).

Also, the meaningful and useful question to ask is not, "How do we optimize control, regardless of environmental (including weather) conditions?", but "How do we optimize control for given expected average conditions in the field?". As explained in chapter 6, we seek an optimum bounded by constraints, and mathematical model-building and manipulation is the only precise, accurate means by which we can determine this optimum.

As Scriven (1959) has clearly pointed out, the fact that a theory is of little use in prediction, due to the inherent unpredictability of independent variables, does not mean that it is useless for explanation. It would have been impossible to predict that dinosaurs would become extinct without knowing that the climate would change. Yet this does not invalidate the theory that evolution is controlled by natural selection. Similarly, a theory can explain why one parasite is usually better than another in a particular environment, without predicting exactly what the relative fates of the two parasites will be in any particular year.

A third criticism that can be levelled against the use of mathematical models in biology is that it is difficult or impossible to determine which model best explains a particular process. For example, Watt (1959a) and Holling (1959) have both proposed reasonable explanations as to why the curve for numbers of hosts or prey attacked bends over as the number vulnerable to attack is increased. However, all such conflicts between models can be resolved by collecting more, and different kinds of data. In many such cases it will probably turn out that the final model contains elements of two or more of the models which preceded it.

ACKNOWLEDGEMENTS

I am indebted to the late Mr. C. Reimer, Dr. P. Robinson and Mr. N. Shklov who read sections of an earlier draft of this paper, and to Mrs. P. M. Morse, with whom I have discussed parts of the paper. However, any errors of omission or commission are the sole responsibility of the author.

REFERENCES

Anderson, R. L., and T. A. Bancroft. 1952. Statistical theory in research. McGraw-Hill, New York.

Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University Chicago Press, Chicago.

Arrow, K. J., L. Hurwicz and H. Uzawa. 1958. Studies in linear and non-linear programming. Stanford U. Press.

Atwal, A. S. 1955. Influence of temperature, photoperiod, and food on the speed of development, longevity, fecundity, and other qualities of the diamond-back moth *Plutella maculipennis* (Curtis) (Lepidoptera: Tineidae). *Aust. J. Zool.* 3: 185-221.

Bailey, N. T. J. 1950. A simple stochastic epidemic. *Biometrika* 37: 193-202.

Bailey, N. T. J. 1957. The mathematical theory of epidemics. Hafner, New York.

Bailey, V. A. 1931. The interaction between hosts and parasites. *Quart. J. Math.* 2: 68-77.

Bancroft, T. A., and T. A. Brindley. 1958. Methods for estimation of size of corn borer populations. *Proc. X Int. Congr. Entomology* 2: 1003-1014.

Baranov, F. I. 1918. On the question of the biological basis of fisheries. *Nauch. issledov. iktiol. inst. izv.* 1: 81-128.

Bartlett, M. S. 1949a. Some evolutionary and stochastic processes. *J. Roy. Stat. Soc. Ser.* B 11: 211-229.

Bartlett, M. S. 1949b. Fitting a straight line when both variables are subject to error. *Biometrics* 5: 207-212.

Bartlett, M. S. 1956. An introduction to stochastic processes with special reference to methods and applications. Cambridge U. Press.

Bartlett, M. S. 1957. On theoretical models for competitive and predatory biological systems. *Biometrika* 44: 27-42.

Bellman, Richard. 1957. Dynamic programming. Princeton U. Press, Princeton.

Bennett, C. A., and N. L. Franklin. 1954. Statistical analysis in chemistry and the chemical industry. John Wiley and Sons, New York.

Berkson, J. 1958. Smoking and lung cancer: some observations on two recent reports. *J. Amer. Stat. Ass.* 53: 28-38.

Beverton, R. J. H., and S. J. Holt. 1957. On the dynamics of exploited fish populations. *Fishery Investigations*, Ser. II, Vol. XIX. Her Majesty's Stationery Office, London.

Birch, L. C. 1953. Experimental background to the study of the distribution and abundance of insects. I. The influence of temperature, moisture and food on the innate capacity for increase of three grain beetles. *Ecology* 34: 698-711.

Bishop, Y. M. M. 1959. Errors in estimates of mortality obtained from virtual populations. *J. Fish. Res. Bd. Canada* 16: 73-90.

Box, G. E. P. 1954. The exploration and exploitation of response surfaces: some general considerations and examples. *Biometrics* 10: 16-60.

Box, G. E. P., and G. A. Coutie. 1956. Application of digital computers in the exploration of functional relationships. The Proceedings of the institution of electrical engineers Vol. 103, Part B Supplement No. 1.

Box, G. E. P., and K. B. Wilson. 1951. On the experimental attainment of optimum conditions. *J. R. Stat. Soc.* B 13: 1-38.

Box, G. E. P., and P. V. Youle. 1955. The exploration and exploitation of response surfaces: An example of the link between the fitted surface and the basic mechanism of the system. *Biometrics* 11: 287-323.

Brown, A. W. A. 1958. Insecticide resistance in arthropods. World Health Organization, Geneva.

Burnett, Thomas. 1954. Influences of natural temperatures and controlled host densities on oviposition of an insect parasite. *Physiological Zoology* 3: 239-248.

Burnett, Thomas. 1956. Effects of natural temperatures on oviposition of various numbers of an insect parasite (Hymenoptera, Chalcididae, Tenthredinidae). *Annals Entomological Society of Amer.* 1: 55-59.

Chiang, C. L. 1954. Competition and other interactions between species. *Statistics and Mathematics in Biology*: 197-215. Iowa State College Press, Ames.

Cragg, J. B., and N. W. Pirie. 1955. The numbers of man and animals. Oliver and Boyd, Edinburgh and London.

d'Ancona, U. 1942. La lotta per l'esistenza. Giulio Einaudi, Torino, Italy.

Davis, D. S. 1955. Nomography and empirical equations. Reinhold, New York.

Deming, W. Edwards. 1938. Statistical adjustment of data. Wiley, New York.

Dempster, J. P. 1957. The population dynamics of the Moroccan locust (*Dociostaurus maroccanus* Thunberg) in Cyprus. *Anti-Locust Bulletin* 27.

Dorfman, R., P. A. Samuelson and P. A. Solow. 1958. Linear programming and economic analysis. McGraw-Hill, N.Y.

Dowden, P. B., H. A. Jaynes and V. M. Carolin. 1953. The role of birds in a spruce budworm outbreak in Maine. *J. Econ. Entomol.* 46: 307-312.

Dwight, H. B. 1947. Tables of integrals and other mathematical data. Macmillan, New York.

Elderton, W. P. 1938. Frequency curves and correlation. Cambridge University Press.

El-Sawaf, Saleh Kamel. 1956. Some factors affecting the longevity, oviposition, and rate of development in the Southern Cowpea Weevil *Callosobruchus maculatus* F. (Coleoptera: Bruchidae). *Bull. Soc. Entomologique D'Egypte* 40: 29-95.

Ezekiel, M. 1941. Methods of correlation analysis. (2nd ed.) Wiley, New York.

Feller, W. 1940. On the logistic law of growth and its empirical verifications in biology. *Acta. biotheoretica* 5: 51-66.

Feller, W. 1950. An introduction to probability theory and its applications. Wiley, New York.

Finney, D. J. 1953. An introduction to statistical science in agriculture. Oliver and Boyd, Edinburgh.

Fraser, D. A. S. 1958. Statistics: an introduction. Wiley, New York.

Fry, F. E. J. 1947. University of Toronto Studies. Biological series No. 55. Publications of the Ontario Fisheries Research Laboratory LXVIII. Effects of the environment on animal activity.

Fry, F. E. J., and K. E. F. Watt. 1957. Yields of year classes of the smallmouth bass hatched in the decade of 1940 in Manitoulin Island waters. *Trans. Am. Fish Soc.* 85: 135-143.

Gass, S. I. 1958. Linear programming: methods and applications. McGraw-Hill, New York.

Gause, G. F. 1934. The struggle for existence. Williams and Wilkins, Baltimore.

Gheury de Bray, M. E. J. 1928. Exponentials made easy. Macmillan and Co., Ltd., London.

Ghiselin, B. 1952. The creative process. U. California Press, Berkeley.

Goulden, C. H. 1952. Methods of statistical analysis. 2nd ed. John Wiley and Sons, New York.

Grumman, H. R. 1936. Some notes on exponential analysis. *Ann. Math. Stat.* 7: 133-144.

Hadamard, J. 1945. The psychology of invention in the mathematical field. Princeton U. Press, Princeton, N.J.

Hader, R. J., M. E. Harward, D. D. Mason and D. P. Moore. 1957. An investigation of some of the relationships between copper, iron and molybdenum in the growth and nutrition of lettuce: I. Experimental design and statistical methods for characterizing the response surface. *Soil Science Soc. Amer. Proc.* 21: 59-64.

Hald, A. 1952. Statistical theory with engineering applications. John Wiley, New York.

Hamaker, H. C. 1955. Experimental design in industry. *Biometrics* 11: 257-286.

Hansen, M. H., W. N. Hurwitz and W. G. Madow. 1953. Sample survey methods and theory. Vol. 1: Methods and applications. Wiley and Sons, New York.

Hartley, H. O. 1948. The estimation of non-linear parameters by internal least squares. *Biometrika* 35: 32-45.

Holling, C. S. 1959. Some characteristics of simple types of predation and parasitism. *Can. Ent.* 91: 385-398.

Hotelling, H. 1927. Differential equations subject to error, and population estimates. *J. Amer. Stat. Ass.* 22: 283-314.

Hueck, H. J., D. J. Kuenen, P. J. DenBoer and E. Jaeger-Draafsel. 1952. The increase of egg production of the fruit tree red spider mite (*Metatetranychus ulmi* Koch) under influence of DDT. *Physiologia Comparata et Oecologia* 2: 371-377.

Ivlev, V. S. 1945. The biological productivity of waters. *Uspekhi Sovremennoi Biologii* 19: 98-120.

Jan'sch, E. 1927. Das exponential gesetz als grundlage einer vergleichenden biologie. Julius Springer, Berlin.

Janisch, E. 1928. Die lebens-und entwicklungsdauer der insekten als temperaturfunktion. *Z. wiss. zool.* 132: 176-186.

Janisch, E. 1932. The influence of temperature on the life-history of insects. *Trans. Ent. Soc.* London 80: 137-168.

Johnson, F. H., H. Eyring and M. J. Polissar. 1954. The kinetic basis of molecular biology. Wiley, New York.

Johnson, L. H. 1952. Nomography and empirical equations. John Wiley and Sons, New York.

Kemeny, J. G., J. L. Snell and G. L. Thompson. 1957. Introduction to finite mathematics. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Kendall, D. G. 1949. Stochastic processes and population growth. *J. Roy. Stat. Soc.* B 11: 230-265.

Knutson, H. 1955. Modifications in fecundity and life span of *Drosophila melanogaster* Meigen following sublethal exposure to an insecticide. *Annals Ent. Soc. Amer.* 48: 35-39.

Koopmans, T. C. 1951. Activity analysis of production and allocation. Wiley and Sons, N.Y.

Kostitzin, V. A. 1937. Biologie mathematique. Collection Armand Colin.

Kramer, C. Y. 1957. Simplified computations for multiple regression. *Ind. Qual. Cont.* 13: 1-4.

Kuenen, D. J. 1958. Influence of sublethal doses of DDT upon the multiplication race of *Sitophilus granarius* (Coleopt. Curculion dae). *Ent. Exp. Appl.* 1: 147-1552.

Kuhn, H. W., and A. W. Tucker. 1951. Non-linear programming. U. California Press, Berkeley.

LeRoux, E. J., and C. Reimer. 1959. Sampling variation of population and mortality factors of immature stages of the eye-spotted budmoth, *Spilonota ocellana* (D. and S.) (Lepidoptera: Olethreutidae), and the pistol casebearer, *Coleophora malivorella* Riley (Lepidoptera: Coleophoridae), on apple in Quebec. *Can. Ent.* 91: 428-449.

Leslie, P. H. 1946. Population theories: A review of La lotta per l'esistenza by U. d'Ancona. *J. An. Ecol.* 15: 107.

Leslie, P. H. 1957. An analysis of the data from some experiments carried out by Gause with populations of the Protozoa, *Paramecium aurelia* and *Paramecium caudatum*. *Biometrika* 44: 314-327.

Leslie, P. H. 1958. A stochastic model for studying the properties of certain biological systems by numerical methods. *Biometrika* 45: 16-31.

Lindley, D. V. 1947. Regression lines and the linear functional relationship. *Roy. Stat. Soc.* 9: 218-244.

Lipka, Joseph. 1918. Graphical and mechanical computation. John Wiley and Sons, N.Y.

Lotka, Alfred J. 1925. Elements of physical biology. Williams and Wilkins Co., Baltimore.

Matsazawa, H., H. Okamoto and Y. Mizamoto. 1957. Application of Pradhan's formula to the data of development of the common cabbage butterfly, *Pieris rapae crucivora*, and its parasite, *Apanteles glomeratus*. *Kontyu* 25: 89-93.

Menusan, H. 1934. Effects of temperature and humidity on the life processes of the bean weevil, *Bruchus obtectus* Say. *Ann. Ent. Soc. Amer.* 27: 515-526.

Meyer, H. A. 1956. Symposium on Monte Carlo methods. Wiley, New York.

Miller, C. A. 1955. A technique for assessing spruce budworm larval mortality caused by parasites. *Can. J. Zool.* 33: 5-17.

Miller, C. A. 1957. A technique for estimating the fecundity of natural populations of the spruce budworm. *Can. J. of Zoology* 35: 1-13.

Miller, C. A. 1958. The measurement of spruce budworm populations and mortality during the first and second larval instars. *Can. J. Zool.* 36: 409-422.

Moore, D. P., M. E. Harward, D. D. Mason, R. J. Hader, W. L. Lott and W. A. Jackson. 1957. An investigation of some of the relationships between copper, iron, and molybdenum in the growth and nutrition of lettuce: II. Response surfaces of growth and accumulations of Cu and Fe. *Soil Science Soc. Amer. Proc.* 21: 65-74.

Moran, P. A. P. 1959. The theory of some genetical effects of population subdivision. *Aust. J. Biol. Sci.* 12: 110-116.

Morris, R. F. 1955. The development of sampling techniques for forest insect defoliators, with particular reference to the spruce budworm. *Can. J. Zool.* 33: 225-294.

Morris, R. F. 1957. The interpretation of mortality data in studies on population dynamics. *Can. Ent.* 2: 49-69.

Morris, R. F. 1959. Single-factor analysis in population dynamics. *Ecology*. 40: 580-588.

Morris, R. F. 1960. Sampling insect populations. *Ann. Rev. Ent.* 5: 243-264.

Morris, R. F., and C. A. Miller. 1954. The development of life tables for the spruce budworm. *Can. J. Zool.* 32: 283-301.

Morris, R. F., C. A. Miller, D. O. Greenbank and D. G. Mott. 1958. The population dynamics of the spruce budworm in Eastern Canada. *Proc. X Internat. Congr. Ent.* 4: 137-149.

Moser, H. 1958. The dynamics of bacterial populations maintained in the chemostat. *Carnegie Inst. Washington Publ.* 614, Washington, D.C.

Nair, K. R. 1954. The fitting of growth curves. *Statistics and mathematics in biology*: 119-132. Iowa State Coll. Press, Ames.

Neyman, J., and E. L. Scott. 1959. Stochastic models of population dynamics. *Science* 130: 303-308.

Nicholson, A. J., and V. A. Bailey. 1935. The balance of animal populations — part 1. *Proc. Zool. Soc.* London: 551-598.

Nielsen, K. L. 1956. Methods in numerical analysis. Macmillan, New York.

Paloheimo, J. E. 1958. A method of estimating natural and fishing mortalities. *J. Fish. Res. Bd. Canada* 15: 749-758.

Parke, N. G. 1958. Guide to the literature of mathematics and physics, including related works on engineering science. Dover Publications, New York.

Payne, N. M. 1933. The differential effect of environmental factors upon *Microbracon hebetor* Say (Hymenoptera: Braconidae) and its host *Ephestia Kuhniella* Zeller (Lepidoptera: Pyralidae). *Biol. Bull. 65*: 187-205.

Poincare, Henri. 1910. The future of mathematics. *Ann. Rept. Smithsonian Institutes* 123-140.

Pradhan, S. 1945. Insect population studies II. Rate of insect-development under variable temperature of the field. *Proc. Nat. Inst. Sci. India* 11: 74-80.

Pradhan, S. 1946a. Insect population studies III. Idea of biograph and biometer. *Proc. Nat. Inst. Sci. India* 12: 301-314.

Pradhan, S. 1946b. Insect population studies IV. Dynamics of temperature effect on insect development. *Proc. Nat. Inst. Sci. India* 12: 385-404.

Pradhan, S., and H. M. L. Srivastata. 1956. Increased activity of insects associated with increased pick-up of insecticide at higher temperatures. *Indian J. Ent.* 18: 78-79.

Pringle, J. W. S. 1957. Insect flight. Cambridge University Press.

Pruthi, H. S. 1956a. Critical studies on insecticides I. Effect of temperature on the toxicity and comparative efficacy of some modern anti-locust insecticides. *Indian J. Ent.* 18: 273-295.

Pruthi, H. S. 1956b. Critical studies on insecticides II. Toxicity of malathion under different conditions of temperature and humidity. *Indian J. Ent.* 18: 408-426.

Quenouille, M. H. 1950. Introductory statistics. Butterworth-Springer Ltd., London.

Quenouille, M. H. 1952. Associated measurements. Butterworths Scientific Publications, London.

The Rand Corporation. 1955. A million random digits with 100,000 normal deviates. The Free Press, Glencoe, Illinois.

Rashevsky, N. 1948. Mathematical biophysics. U. Chicago Press.

Ricker, W. E. 1958. Handbook of computations for biological statistics of fish populations. Bulletin 119, *Fisheries Research Board of Canada.*

Riemschneider, R. 1954. Configurations et action de certains insecticides. Examen stereochimique et toxicologique d'analogues du DDT. *Chimie and industrie* 72: 261-270, 435-443.

Roeder, Kenneth D. 1953. Insect physiology. John Wiley, New York.

Running, T. R. 1917. Empirical formulas. John Wiley, New York.

Salt, R. W. 1950. Time as a factor in the freezing of undercooled insects. *Can. J. Res. D.* 28: 285-291.
Salt, R. W., and H. G. James. 1947. Low temperature as a factor in the mortality of eggs of *Mantis religiosa*. *Can. Ent.* 79: 33-36.
Schaefer, M. B. 1957. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical Pacific Ocean. *Inter-American Tropical Tuna Commission Bulletin* 11: 247-285.
Schmalhausen, I. I. 1949. Factors of evolution. Blakiston, Philadelphia.
Scriven, Michael. 1959. Explanation and predictions in evolutionary theory. *Science* 130: 477-482.
Skellam, J. G. 1951. Random dispersal in theoretical populations. *Biometrika* 38: 196-218.
Smith, F. D. 1952. Experimental methods in population dynamics: A critique. *Ecology* 33: 441-450.
Smith, H. S., and P. DeBach. 1942. The measurement of the effect of entomophagous insects on population densities of the host. *J. Econ. Ent.* 35: 845-849.
Smith, R. W. 1958. Parasites of nymphal and adult grasshoppers (Orthoptera: Acrididae) in Western Canada. *Can. J. Zool.* 36: 217-262.
Snedecor, G. W. 1956. Statistical methods. Iowa State Coll. Press, Ames, Iowa.
Sotavalta, O. 1947. The flight-tone (wing-stroke frequency) of insects. *Acta Ent. Fenn.* 4: 1-117.
Sotavalta, Olavi. 1952. Flight-tone and wing-stroke frequency of insects and the dynamics of insect flight. *Nature* 170: 1057-1058.
Sotavalta, O. 1954. On the thoracic temperature of insects in flight. *Ann. Zool. Soc.* 'Vanamo' 16: 1-22.
Spencer, G. J. 1958. The natural control complex affecting grasshoppers in the dry belt of British Columbia. *Proc. X Internal. Congr. Ent.* 4: 497-502.
Stanley, J. 1932. A mathematical theory of the growth of populations of the flour beetle, *Tribolium confusum* Duval. *Can. J. Res.* 6: 632-671.
Steinhaus, Edward A. 1954. The effects of disease on insect populations. *Hilgardia* 23: 197-261.
Stevens, W. L. 1951. Asymptotic regression. *Biometrics* 7: 247-267.
Tahori, A. S. 1955. Diaryl-trifluoromethyl-carbinols as Synergists for DDT against DDT-resistant house flies. *J. Econ. Ent.* 48: 638-642.
Thompson, W. R. 1924. La theorie mathematique de l'action des parasites entomophages et le facteur du hasard. *Ann. Fac. Sci.* Marseille 2: 69-89.
Thompson, W. R. 1928. A contribution to the study of biological control and parasite introduction in continental areas. *Parasitology* 20: 90-112.
Thompson, W. R. 1939. Biological control and the theories of the interactions of populations. *Parasitology* 31: 299-388.
Tukey, J. W. 1949. One degree of freedom for non-additivity. *Biometrics* 5: 232-242.
Ullyett, G. C. 1953. Biomathematics and insect population problems. A critical review. *Memoirs Ent. Soc.* Southern Africa 2: 1-89.
Vajda, S. 1956. The theory of games and linear programming. John Wiley, New York.
Vazsonyi, A. 1958. Scientific programming in business and industry. John Wiley and Sons, New York.
Volterra, V. 1926. Variazione e fluttuaziomi del numero d'individui in specie animali conviventi. *Mem. Acad.* Lincei 6: 31-113.
Wallis, W. A., and H. V. Roberts. 1956. Statistics: a new approach. The Free Press, Glencoe, Illinois.
Walsh, J. W. T. 1919. The resolution of a curve into a number of exponential components. *Proc. Phys. Soc.* London 32: 26-30.
Watt, K. E. F. 1955. Studies on population productivity. I. Three approaches to the optimum yield problem in populations of *Tribolium confusum*. *Ecol. Monogr.* 25: 269-290.
Watt, K. E. F. 1956. The choice and solution of mathematical models for predicting and maximizing the yield of a fishery. *J. Fish. Res. Bd., Can.* 13: 613-645.
Watt, K. E. F. 1959a. A mathematical model for the effect of densities of attacked and attacking species on the number attacked. *Can. Ent.* 91: 129-144.
Watt, K. E. F. 1959b. Studies on population productivity. II. Factors governing productivity in a population of smallmouth bass. *Ecol. Monogr.* 29: 367-392.
Watt, K. E. F. 1960. The effect of population density on fecundity in insects. *Can. Ent.* 92.
Weiss, P., and J. L. Kavanau. 1957. A model of growth and growth control in mathematical terms. *J. Gen. Physiology* 41: 1-47.

Wigglesworth, V. B. 1950. The principles of insect physiology. Methuen and Co., London.
Will, H. S. 1936. On a general solution for the parameters of any function with application to the theory of organic growth. *Ann. Math. Stat.* 7: 165-190.
Wilson, E. B., and R. R. Puffer. 1933. Least squares and laws of population growth. *Proc. Amer. Acad. Arts and Sciences* 68: 285-382.
Wolcott, G. N. 1958. The evanescence of perfect biological control. *Proc. X Int. Congr. Ent.* 4: 511-513.
Worth ng, A. G., and J. Geffner. 1943. Treatment of experimental data. John Wiley and Sons, Inc., N.Y.
Wright, R. H. 1958. The olfactory guidance of flying insects. *Can. Ent.* 90: 81-89.
Wright, S. 1937. The distribution of gene frequencies in populations. *Proc. Nat. Acad. Sc.* 23: 307-320.

The Runge Press Ltd.
Mailed February 24th, 1961.